USING MAXIMUM ENTROPY TO MODEL THE DISTRIBUTION OF AN ENDANGERED, ENDEMIC CRAYFISH *CAMBARUS HARTI* (HOBBS)

James Kiourtsis

COLUMBUS STATE UNIVERSITY

USING MAXIMUM ENTROPY TO MODEL THE DISTRIBUTION OF AN ENDANGERED, ENDEMIC CRAYFISH *CAMBARUS HARTI* (HOBBS)

A THESIS SUBMITTED TO

THE COLLEGE OF LETTERS AND SCIENCE

IN PARTIAL FULFILLMENT OF

THE REQUIREMENTS FOR THE DEGREE OF

MASTER OF SCIENCE

DEPARTMENT OF EARTH AND SPACE SCIENCES

BY

JAMES KIOURTSIS

COLUMBUS, GEORGIA

2018

ABSTRACT

Cambarus harti is a state-listed endangered, endemic crayfish found only in three counties in mid-west Georgia. Several studies have attempted to characterize the biology and ecology of this crayfish, however data regarding the distribution of this rare, endemic crayfish remains limited. The International Union for Conservation of nature stated that in order to create an effective conservation plan, the known distribution must be expanded. Species distribution models are a cost-effective way to identify locations that have similar habitat characteristics to those with known populations. One species distribution model, Maximum Entropy (MaxEnt), is the preferred approach when modeling species, like *C. harti*, that only have a few known locations. I used MaxEnt to create a predictive, spatial model for *C. harti.* The MaxEnt model was developed using 14 *C. harti* occurrence locations and five environmental layers (distance to water, soils, geology, landcover, and slope) for six counties in West Central Georgia. Using a 2km buffer for background points the model produced a receiver operating characteristic curve (ROC) with an area under the curve (AUC) value of 0.97. The high AUC value correlates with the high discriminatory power of the model. The five environmental layers were weighted differently starting with the most important; distance to water (35.4%), soil (29.1%), landcover (14.8%), geology (14.3%), and slope (6.3%). The model's results covered 6110 km^2 in Georgia with probabilities of *C. harti* occurrence ranging from: 0%-100% [(0%-10%) 4432 km^2, (10%-20%) 622km^2, (20%-30%) 371 km^2, (30%-40%) 214 km^2, (40%-50%) 150 km^2, (50%-60%) 137 km^2, (60%-70%) 122 km^2, (70%-80%) 30 km^2, (80%-90%) 30 km^2, (90%-100%) 2 km^2]. The MaxEnt model was evaluated through two different ground truthing methods. The first approach examined the model's overall accuracy by randomly sampling for crayfish at 30 sites across three model

predicted probability ranges (0%-20%, 40%-60%, 80%-100%). The second method evaluated the model output at finer resolutions by comparing probabilities of known *C. harti* locations to sites within 183m of known locations but without crayfish. The first approach yielded no verified *C. harti* locations within any of the sampling brackets. The second method confirmed that the model was ineffective at identifying *C. harti* habitat on large spatial scales (i.e. locally).

Review of the environmental data layers used to create the model uncovered errors in the underlying data. For example, the USGS National Hydrology Dataset was a large source of error, with many streams improperly mapped. This data set was used to create the distance to water grid. It is clear that data resolution, accuracy and resolution have not advanced to the point where these models can justifiably be used to map the potential habitat of this endemic burrowing crayfish. *Cambarus harti* is likely just one of many species this model is inadequate for; models for amphibians and other species that rely on ground water or surface water depicted by the USGS National Hydrology Dataset would lack adequate data. A high resolution (10m) groundwater layer needs to be obtained in order to more accurately model burrowing crayfish habitat.

ACKNOWLEDGEMENTS

I would like to thank the members of my committee for all the help and engorgement they provided along the way. Dr. Troy Keller, thank you for the opportunity to work alongside you over the years and for providing me with impromptu guidance when I was struggling. Dr. Chester Figiel I would like to thank you for sharing with me some of your hands on experience in the field and allowing me to use some of your most recently found *C. harti* locations in my thesis. I would like to thank Dr. Clifton Ruehl for his guidance on field sampling bias as well as statistical support. Also I would like to give a big thank you to Zach Sertell for coming out into the field with me and helping me search for this crayfish no matter how hazardous the sampling location may have been.

Finally, I would like to thank my family for providing me with an unparalleled amount of support at all hours of the day no matter how many times I called home they always picked up.

LIST OF TABLES

LIST OF FIGURES

Introduction

Biodiversity, the diversity of genes, species, and habitats within ecosystems, is arguably the most important driver of ecosystem functions (Zavaleta, 2010). A healthy diverse ecosystem is able to provide services such as clean air and water along with aiding the removal of pollutants (Schläpfer, 1999). One of the leading threats to biodiversity is the expanding anthropogenic degradation to the environment. Threats such as sediment loading, pollution, and sprawling cities destroy natural habitat (Singh, 2002). These anthropogenic effects have led to an increase in species extinction rates around the globe, and placed other species in peril (Singh, 2002). Often overlooked are the less charismatic organisms, such as: worms, ants, microbes, and crayfish, among other macroinvertebrates. These small organisms have been shown to play a critical role in ecosystems where they process organic matter that facilitates nutrient cycling (Covich et al., 1999). Despite their small size, these invertebrates (e.g. crayfish, snails and nymphs) are critically important for maintaining healthy aquatic ecosystems because they often occur in high densities (Wallace & Webster, 1996).

On a global scale freshwater ecosystems have suffered the largest percent loss of biodiversity, making a clear case for the conservation of freshwater habitats (Richman et al., 2015). Even though freshwater ecosystems only occupy around 1% of the Earth's surface area they support around 10% of all known species (Strayer & Dudgeon, 2010). Among freshwater species, crayfish are highly imperiled. Nearly 32% of the approximately 590 freshwater species worldwide are at risk of extinction (Richman et al., 2015).

In the U.S., only 52% of the 363 identified crayfish species are listed as stable; the other 48% are threatened, endangered or possibly extinct (Taylor et al., 2007). The southeastern United States is a hotspot of freshwater biodiversity (Master et al. 1998, Georgia DNR - Wildlife Resources Division, 2017) particularly crayfish diversity. Georgia is home to 68 native crayfishes and 3 non-native species (Skelton, 2010, Crayfish of U.S., 2017). Of the 68 native species, a third of Georgia's crayfish species are at risk of extinction (Skelton, 2010, Department Of Natural Resources Division, 2017) . Some species are at risk due to their low population sizes and their limited range (Skelton, 2010, Department Of Natural Resources Division, 2017).

Cambarus harti, the Piedmont Blue Burrower, is state-listed endangered, endemic crayfish species with a distribution limited to areas within and near Meriwether County in West Central Georgia (Keller et al., 2011). *Cambarus harti* is often found in forested wetland habitats with shallow groundwater (Keller et al., 2011, Helms et al., 2013, Gilmer, 2014). These scientific studies are based on a few observations that had relatively small population sizes and a limited number of locations. Studies of other primary burrowing crayfish suggest that crayfish must be able to connect with groundwater (Hobbs, 1981, Skelton, 2010, Keller et al., 2011). Crayfish have gills in their carapace that must be damp in order for them to respire (Tarr, 1884, Hasiotis, 1993, Skelton et al., 2002, Loughman, 2010). Many primary burrowers live near streams (Tarr, 1884, Hobbs, 1981,), presumably because the clay soils and impermeable geologic layers elevate groundwater near the ground surface. Hobbs, (1981) observed that *C. harti* was found in locations where the terrestrial habitats transitioned into flood plains.

Hobbs (1981) originally described this species as a primary burrower that inhabits underground tunnels connected to the groundwater. This species was commonly found near

springs and seeps (Hobbs, 1981). *Cambarus harti* excavates complex sets of tunnels. Some tunnels run horizontal into chambers while others are oriented vertical. Hobbs (1981) hypothesized that the chambers were developed to provide refuges during fluctuations in the groundwater. The burrow openings are often marked with chimneys that are typically 10 to 15cm in height (Helms et al, 2013). Hobbs (1981) also noted that *C. harti* retreated down to the deepest chamber when an individual's burrow was being excavated. This behavior made the retrieval of the species particularly difficult (Hobbs, 1981). *Cambarus harti* was described as blue in color. The fourth pair of legs of the crayfish, pereiopod, include simple acute hooked ends (Hobbs, 1981). Hobbs (1981) describes their first set of pleopods, as leg like features attached to the abdomen, extending to the third set of pereiopods. He also noted that both pleopods meet flat against each other with acute tips and there is no sign of a notch on the pleopod on the ventral side.

The State of Georgia, as well as the IUCN, list *C. harti* as endangered due to its narrow range and small population size (Cordeiro et al., 2010, Skelton, 2010). These evaluations should be considered preliminary, because they are based on limited, and in some cases, historical data. For example, the IUCN's information is limited to 2 populations and 16 specimens (Cordeiro et al., 2010). In order to determine adequate conservation strategies, the IUCN stresses that scientists must locate more populations. Although several *C. harti* populations have recently been discovered (Skelton, 2002, Keller et al, 2011), conservation planning for *C. harti* depends on the identification of new populations and the collection of additional ecological data.

The development of species distribution models (SDM) has provided conservationists and environmental scientists a suite of new tools that can be used to identify potential habitats for

species, particularly ones with specific niche requirements. SDMs are valuable because they predict the probability of a species' occurrence across the geographic landscape (Phillips et al., 2005). Most SDMs predict the likelihood of a species' occurrence based on a relationship between known locations and user defined habitat related environmental data in the form of spatial layers such as soils, land cover, and climate (Guisan & Thuiller, 2005). There are two main model types. Presence-absence models require both occurrence and absence locations, while presence-only models require known locations only. Although there are a number of different species distribution model approaches (DOMAIN, MARS, GAM, GBM, GLM, etc.), only maximum-entropy modeling (MaxEnt) has proven effective for endemic species with only a few known populations (Wisz et al., 2008).

MaxEnt is a maximum entropy model with thresholds (Wisz et al., 2008) that can predict a species' distribution based on environmental covariates. The model uses prior data, occurrence locations and environmental layers, to determine the constraints (i.e. mean, variance) applied to the model output. Maximum entropy is so named because there are many different models that could fulfill the input constraints. When creating the final model output MaxEnt starts as a perfectly uniform probability distribution in geographic space, then it applies the constraints forcing the model away from this uniform distribution to create the final model (Elith et al., 2011). MaxEnt can be used to analyze categorical and continuous data types. These forms of data can be modeled using; linear, quadratic, product, threshold, hinge, and binary associations (Elith et al., 2011). In order to produce an accurate model, MaxEnt uses (L-1) regularization to improve machine learning (Hastie et al., 2009, Elith et al., 2011). This technique is commonly used when

multiple factors (i.e. environmental layers) describe one data point, it softens the distribution pushing weight onto more explanatory factors (Hastie et al., 2009, Elith et al., 2011).

In order to create an effective MaxEnt model one must understand how to create and revise the model to remove inaccuracies and adjust parameters to assure a good model fit (Phillips et al., 2004, Phillips et al., 2005, Elith et al., 2011). The precision of the model is fully dependent on the resolution of your spatial data layers. MaxEnt uses prior data collected about the species' location and can be used in applications without absence locations. In the past MaxEnt has been successfully applied to rare and endemic species such as the endangered dwarf wedgemussel and endemic birds in temperate forests of Southern Chile (Wilson et al., 2011, Moreno et al., 2011, Campbell & Hilderbrand, 2016). However, to this author's knowledge, it has only been used to model the distribution of two burrowing crayfish species (Rhoden et al., 2017).

MaxEnt, with its published use on burrowing crayfish (Rhoden et al., 2017), seems the appropriate model for *C. harti.* Effective conservation of *C. harti* depends on a more thorough understanding of this species' distribution. My goal is to use MaxEnt to develop a spatial model that predicts potential habitat and can be used to expand the known distribution of this species. An effective SDM would facilitate research about *C. harti* needed for the development of an effective conservation plan. Only with additional data can scientists help protect this endangered species.

Materials and Methods

Presence data and environmental variables

Species distribution models rely on spatially explicit data that could be useful for accurately predicting a species' distribution. The MaxEnt modeling application requires a CSV file containing

the latitude and longitude of known occurrences (i.e, confirmed locations) and an ASCII grid for each environmental layer (e.g. soils). Using ArcGIS 10.4, all environmental rasters were set to the same extent (Fig. 1), cell size (10m), and projection (NAD 1983 Georgia Statewide Lambert) in order to be included in the MaxEnt model. Cell size was manipulated by converting all layers to match the finest resolution (10m). This approach was an attempt to retain all the data instead of aggregating pixels to a larger pixel size resulting in a loss of data. The environmental layers soil, geology, landcover, slope, and distance to water were included in the model (Table 1). Soil, geology, and landcover were included to describe the burrowing habitat while slope, soil, and distance to water could provide indicators of the hydrologic conditions in the area. All of these layers were hypothesized to play a role in the *C. harti's* habitat requirements. Soil, geology, and land cover were originally in vector form and were converted to raster using ArcTools (polygon to raster). The model extent included six counties located in west central Georgia: Harris, Talbot, Upson, Pike, Meriwether, and Troup (Fig. 1). It is important that the model's extent is chosen to fit the potential range that the species could exist, otherwise the model has a high chance of overfitting inadequate locations (Elith et al., 2011). Across the globe there may be many areas that have the same environmental conditions required by a species. However, the range of the species is a key limiting factor that must be accounted for in the model. The occurrence data were configured in MS Excel™ to the specific structure required by the model. Data were compiled from research identifying 13 known locations reported in Keller et al, (2011, Fig. 1).

Table 1. Environmental data layers that were used in the *Cambarus harti* MaxEnt model, including their scale or resolution, description, and source.

Environmental Layer	Description	Source	Unit Type	Resolution or Source scale
Geology	28 geologic structures, Polygon data	Georgia Clearinghouse	Nominal	1:250,000
Soils	soil characteristics, Polygon data	Georgia Clearinghouse	Nominal	1:250,000
Landcover	28 Landcover types across study site, raster data	USGS	Nominal	30m
Slope	Created from a digital elevation model, raster data	USGS	Ratio	10m
Euclidian Distance to Water	Created from the National Hydrology Dataset, depicts distance from water, line and polygon data	USGS	Interval	1:24,000/1:12,000

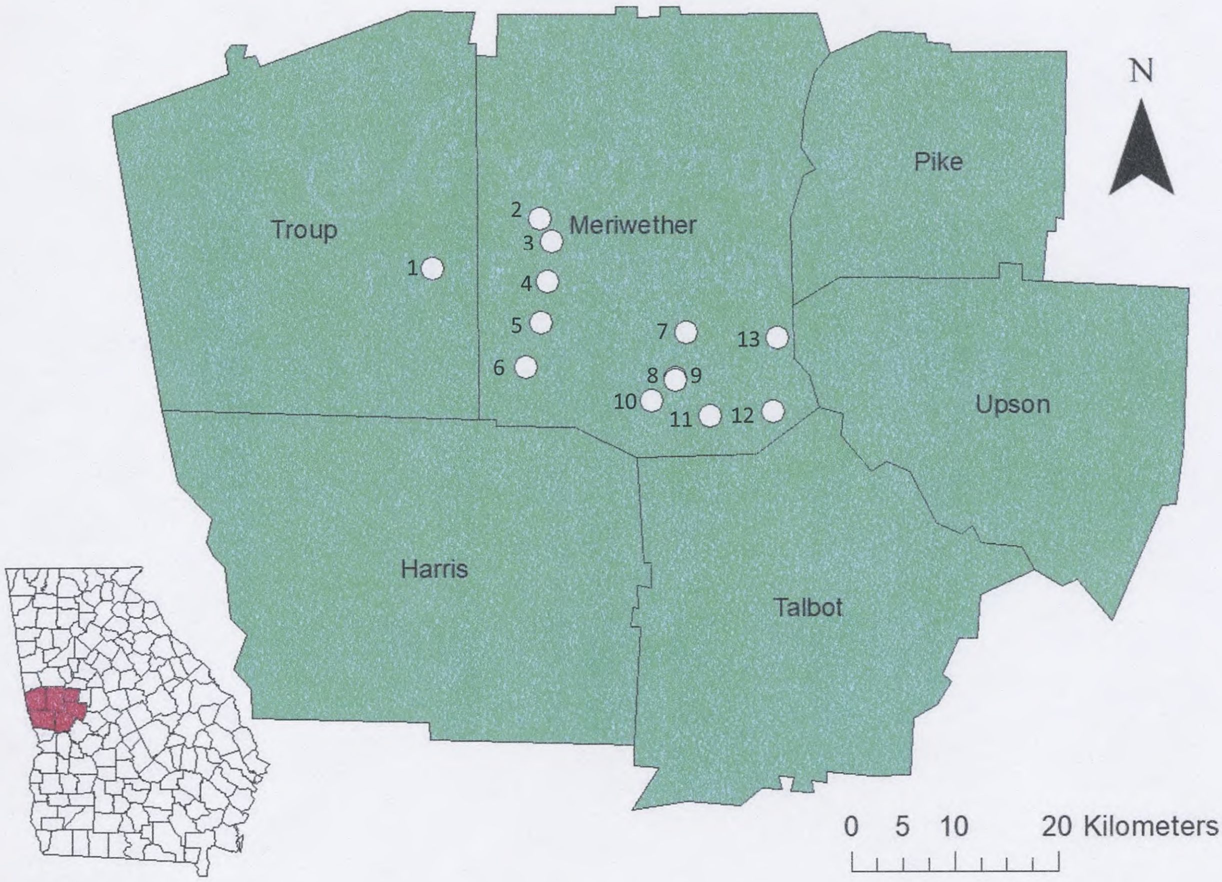


Figure 1. Known *Cambarus harti* locations (n=13) relative to the six counties in Georgia used in the MaxEnt model.

MaxEnt Analysis

The *C. harti* model was developed using MaxEnt software [version 3.3.3k] (Phillips et al, 2011). To help reduce potential sampling bias (Dudik et al., 2005, Phillips, 2008) 10,000 random background points were collected within a 2-km radius of each known location (Peterman et al., 2013). The model was set so the probability output distribution produced would be logistic with potential maximum probability score equal to 1. MaxEnt was also set to create response curves for both the continuous (distance to water, and slope) and categorical (landcover, geology, and soil) data. MaxEnt was set to 5000 iterations of the model and to produce a receiver operating characteristic curve (ROC) including the area under the curve (AUC) value. The output of the model was set to ASCII format so that the resulting predictions could be imported into ArcMap.

Because the model has the potential to include layers that are uninformative, jackknifing was used to evaluate individual layers to determine their importance in the model. The jackknifing algorithm runs the model 11 times removing layers while retaining others to determine the importance each layer has on the overall model. All input layers were retained in the final model (Table 1). ENMeval and ENMTools, packages in R (R Core Team, 2017), were used to ensure that MaxEnt doesn't under or over fit the known distribution (Phillips & Dudik, 2008). Model validity was also evaluated by creating a null model calculated from 13 randomly placed occurrence points (Raes & Steege, 2007) using ENMtools (Warren et al., 2010). The null model was run in MaxEnt the same way my model was and the AUC curves were compared quantitatively when it was finished.

Field Validation: Random Locations

Once the model was fully developed, its accuracy was accessed using field surveys for *C. harti*. Because the model generated a grid that predicted the probability that *C. harti* will be found in each cell, its accuracy could be validated by visiting various pixel locations and searching for *C. harti*. Thirty locations were chosen to be sampled and were split into three groups, 10 with high presence probability (80%-100%), 10 with medium presence probability (40%-60%), and 10 with low presence probability (0%-20%). This technique, modeled after Rhoden et al (2017), facilitates comparison of model performance among levels. While every effort was made to select sampling locations at random, problems with public access required sampling choices to be semi-random. Locations were removed if the land was developed, permission wasn't granted, or if the terrain was inappropriate (ex: lake bottom). Thirty semi-random locations were chosen from a fully random set of 100 locations (Fig. 2).

A field sampling protocol was developed to evaluate model predictions. The area sampled in the field was matched to the cell size of the raster's output (10mx10m). To account for the patchy distribution of *C. harti* populations (Hobbs, 1981), the 8 pixels surrounding the randomly selected sample location were also sampled (Fig. 3). Thus the total sampling area covered 9 cells and a total of 900m^2 at each of the 30 sample locations (270 pixels from the model). This intensive sampling protocol improved the chance of detecting *C. harti* and reduced the potential for false negatives (i.e. missed when present).

It was important that field validation follow a consistent and effective protocol when searching for *C. harti.* For this study, each location (i.e. 9 cells) was examined for crayfish burrows for up to 3 hours. To navigate to the sample location, the latitude and longitude was extracted

from the centroid of the sample pixel using ArcMap; a Garmin GPSmap76CSx (<10m accuracy) was used in the field to navigate to the centroid of the sample location. At the centroid, a picture of the location was taken to accurately depict the habitat characteristics in that area. From this point, an open wheel tape measure was used to plot 8 different flags around the boundary of the sample site (compass directions; N, NE, E, SE, S, SW, W, NW). In order to ensure that the sampling matched all of the appropriate pixels from the model, I placed flags 13.71 meters N, S, E, W of the centroid and 20.12 meters (NE, SE, SW, NW) (Fig. 3). While walking these lines a rake was used to move debris and leaves aside exposing soil to aid in the visual search for crayfish chimneys and burrows. The flagged area was further assessed by walking through the entire sampling site searching for burrows to ensure equivalent sampling effort was allocated for each pixel. If no burrows/chimneys were found at the site, sampling ended prior to the three-hour max sampling period. At each site, I recorded the latitude and longitude, noted the time spent sampling, took digital images with a 12MP camera and measured any crayfish captured. All sample locations were treated similarly regardless of their probability ranking. A Chi-square analysis was run comparing the number of burrows dug at different probability value sites.

When a potential burrow (a hole running vertical into the ground) was located, it was excavated slowly and carefully to ensure that the specimen was not harmed in any way. The hole was dug out using a shovel until groundwater started to fill the passageway, then a plunger pump was used to pump silt-filled water into the burrow in an attempt to force the crayfish to crawl out of the burrow. This approach drops the already low dissolved oxygen levels and, has been used successfully on two preliminary excavations of *C. harti* (Keller et al 2013).

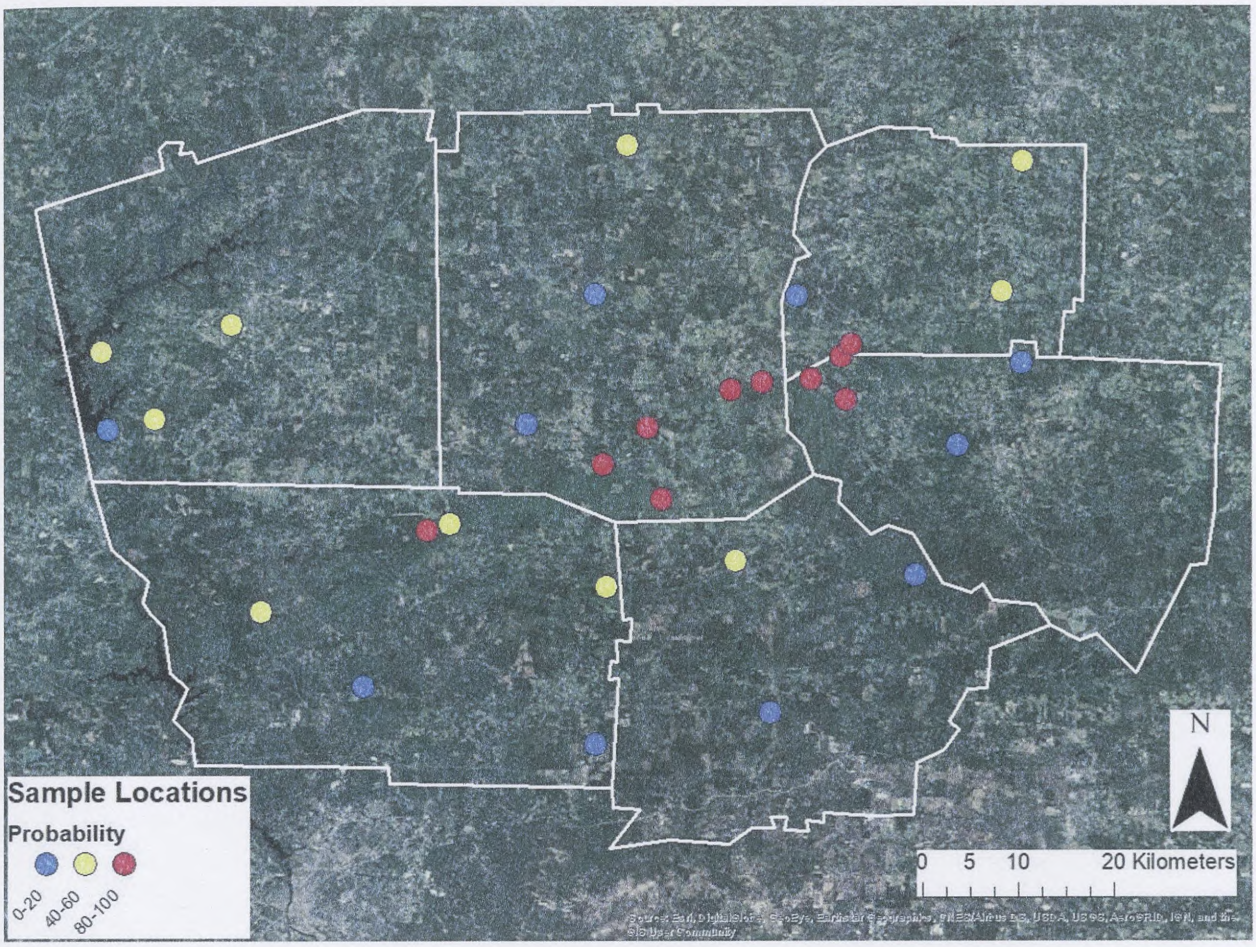


Figure 2. Thirty semi-random locations where sampling of the *Cambarus harti* occurred. The 0%-20% probability locations are shown in blue, the 40%-60% in yellow and the 80%-100% in red. The background is an image from ESRI digital globe with an outline of the six counties used in this study.

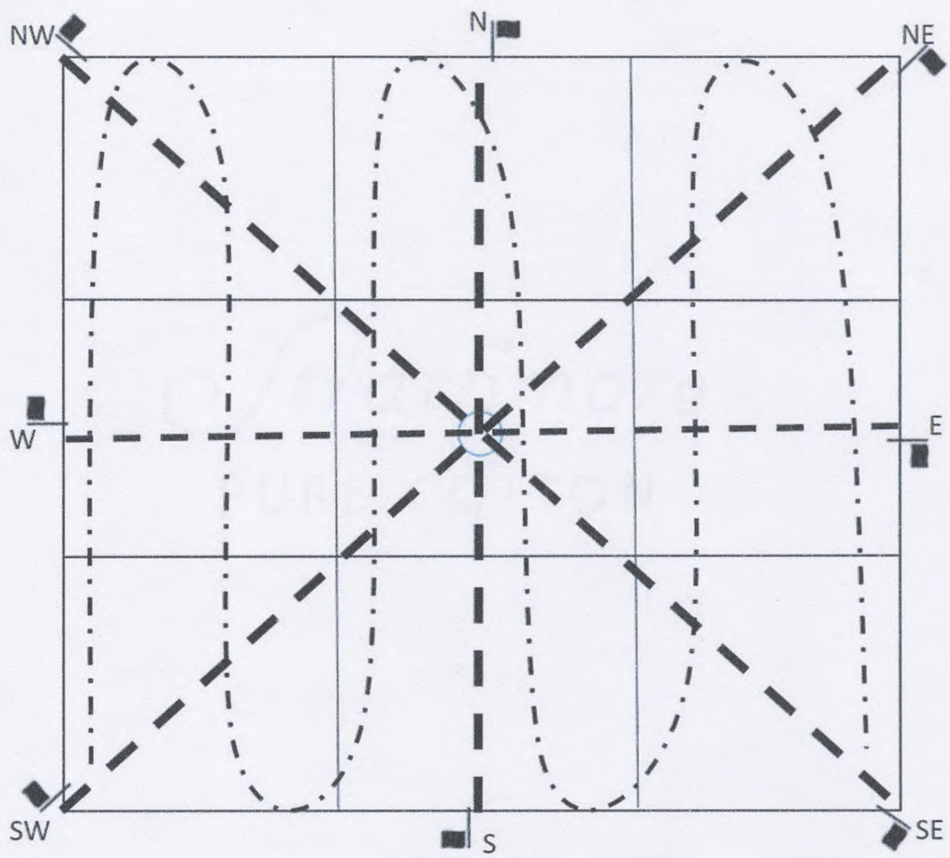


Figure 3. A depiction of the sampling technique used for the field validation of the 30 random sites. The circle represents the starting point, the heavy dashed lines symbolize the path walked with a rake to set flags along the outside of the sampling area. The light dotted line was the area walked searching for burrows.

Field Validation: Known Locations

The second phase of the field validation test compared model predicted probabilities at sites where the crayfish existed with the surrounding areas where it was absent. This phase investigated potential sources of model error at the local scale by determining the accuracy of the model at known *C. harti* locations. Four locations were selected where the species was known to exist (Fig. 4). At each of the sites, I used a Trimble Geo 7X with 1-100cm accuracy to navigate to the known location, then searched the ground for potential burrows in an outward

spiral fashion. Once a burrow was located the latitude and longitude were collected. Because each pixel in the model was 10mX10m I made sure that the burrows were 10m apart. This approach assured that points would fall in different model pixels. If a burrow was less than 10m away this new burrow wasn't plotted.

Following the mapping of the known locations, I searched similar habitats without *C. harti* by selecting an area within 183-m of the known location. To reduce spatial bias, the absence locations were chosen to be equally proximal to the closest surface water source. Had these sites been selected with different distances that would have biased their probability distribution because the model is dependent on distance to water. The same outward spiraling search technique was used to confirm the absence of burrows. To ensure equivalent sampling effort, I searched an area corresponding to nine different model pixels ($752m^2$). This resulted in 72 plotted points, nine presence and nine absences for each of the 4 locations mapped. The latitude and longitude for the locations were collected and stored using the Trimble Geo 7X (accuracy = 1m).

Analysis of the field data was conducted using ArcMap 10.4. After entering the latitude and longitude for presence and absence sites, the data were edited to include the location name (Chandler, Cartwright, FDR Institute, or Warm Springs) and status, (presence or absence of *C. harti)* for all 72 points. The MaxEnt model was uploaded to the ArcMap^{Tm} in ASCII form and converted to a raster using ArcTools (ASCII to Raster). In order to extract values at each pixel, the data in the table needed to be converted to an integer. The data were then converted from a decimal to a percent using raster calculator in ArcTools. The resulting data included many decimal places following the percent making it impossible for ArcMap^{Tm} to produce a raster data

table with so many unique values. ArcTools raster calculator was then used to convert the values to integers. The map was analyzed visually to ensure that each GPS point fell in a different model pixel. Extract by point tool was used to determine the MaxEnt model value at each GPS sample location. These data were exported to MS Excel™ for further analysis.

A statistical comparison was conducted to determine if the model predictions (dependent variable) varied among presence and absence locations (independent variable #1) as well as the properties (independent variable #2) using a two-way ANOVA. Levene's Equality of Variances test was used to test the assumptions of homoscedasticity. A Tukey post-hoc pairwise comparison was used to compare differences between the properties. Each individual property's presence and absence data were compared in SPSS using a difference of least squares means, comparing individual properties as well as comparisons between properties (IBM Corp, 2017). The site labeled the FDR institute was omitted from the ANOVA and least squares means analysis, because all of the probability values were equal to 0%.

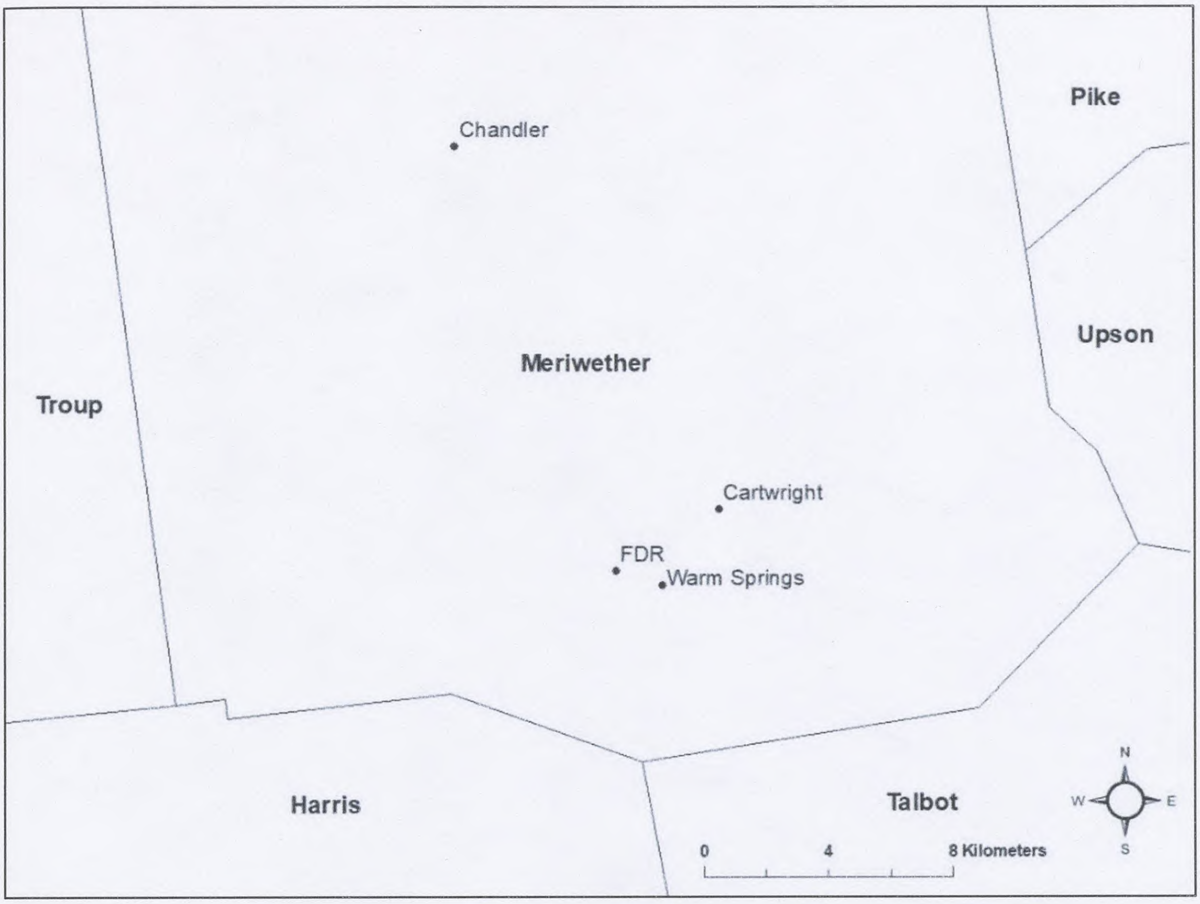


Figure 4. Four locations used for field validation of known locations, as well as the six labled counties used as the extent of this study.

Results

Presence data

The presence data (N=14) for the model was based on historical surveys of *C. harti* (Fig. 1, Keller et al, 2011). The 14 known locations had slopes that ranged from 0-2% grade and were located near surface water (max distance 75m). These locations were recorded mostly in hardwood forests where the underlying geology consisted of mica schist and mica schist/gneiss.

Maxent analysis

The MaxEnt model scored an overall AUC value of 0.957 (Fig. 5). The model weighed the environmental layers as follows: distance to water (35.4%), soil (29.1%), landcover (14.8%),

geology (14.3%), and slope (6.3%). A jackknife analysis showed that each layer played an important role in the overall model AUC (Fig. 6), so all layers were retained however landcover and slope had the lowest regularized gains.

The results showed that probability of occurrence had a strong negative relationship with distance to water and slope (Fig. 7A). Distance to water drops to a probability of almost 0 beyond 305m. Slope also drops to probabilities equal to 0 when slope is greater than 4% (Fig. 7B). Generally the species' known locations were found on, mica schist (Fig. 7D) and hardwood forest (Fig. 7E). The model's spatial extent consisted of 6110 km^2 with probabilities of occurrence ranging from 0%-100% [(0%-10%) 4432km^2, (10%-20%) 622km^2, (20%-30%) 371km^2, (30%-40%) 214km^2, (40%-50%) 150km^2, (50%-60%) 137km^2, (60%-70%) 122km^2, (70%-80%) 30km^2, (80%-90%) 30km^2, (90%-100%) 2km^2] (Fig. 8). As the probability of occurrence increased the amount of predicted area decreased. Relative to the whole study area, high probability habitats (80%-100%) only encompassd 0.5% of the total area. This pattern is small but reflective of a niche species such as *C. harti*.

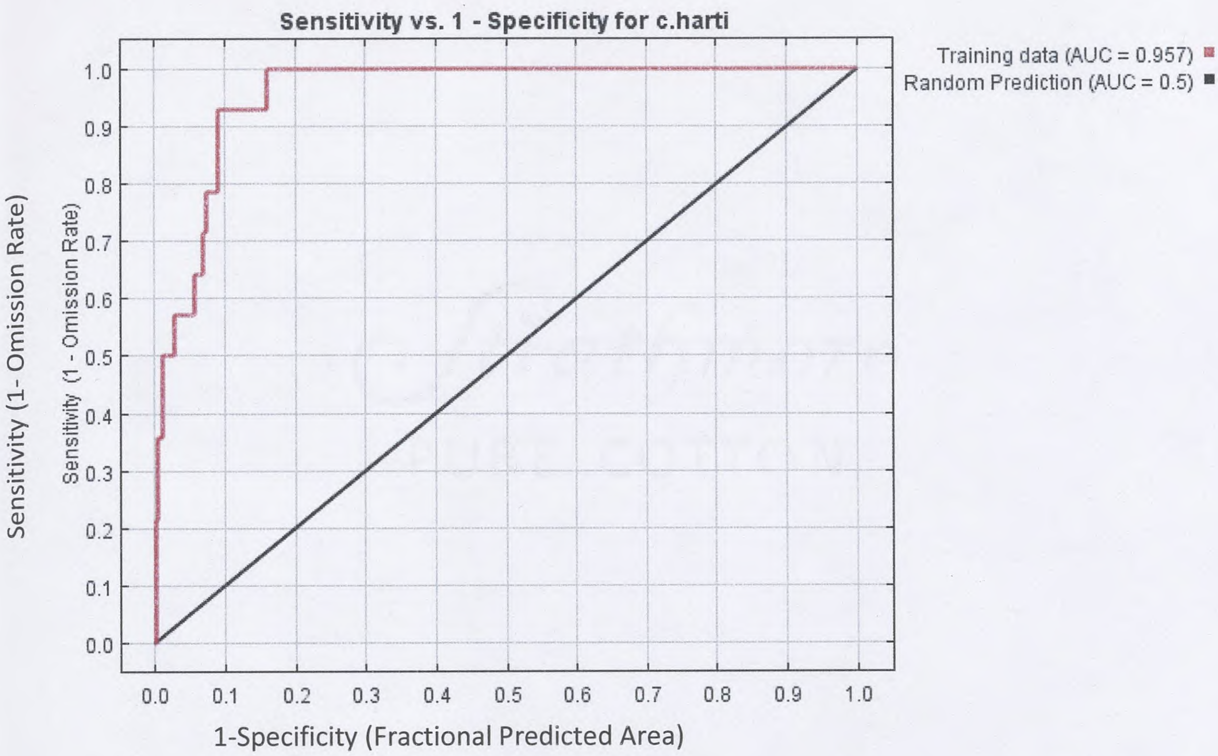


Figure 5. Effects of sample size on the model's predictive capability (i.e. % of known locations). The *Cambarus harti* model scored an AUC (area under the curve) of 0.957. The red line symbolizes an AUC value for the model while the black line represents a random predicted AUC value.

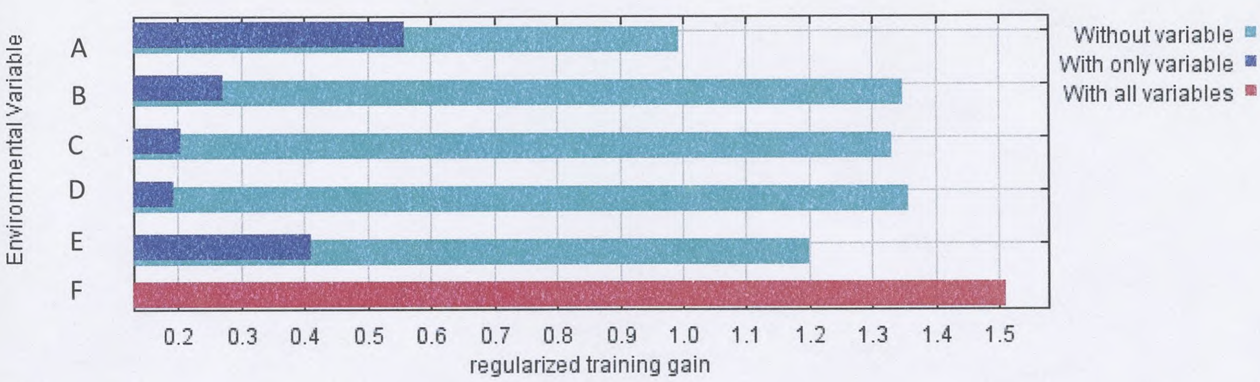


Figure 6. Jackknife of regularized AUC training gain for *Cambarus harti* model for each layer. A) distance to water, B) geology, C) landcover, D) slope, E) soil, and F) all layers combined. The Blue color symbolizes how the model's AUC (area under the curve) will be affected with only that variable present and the turquoise shows AUC values with the removal of only that variable.

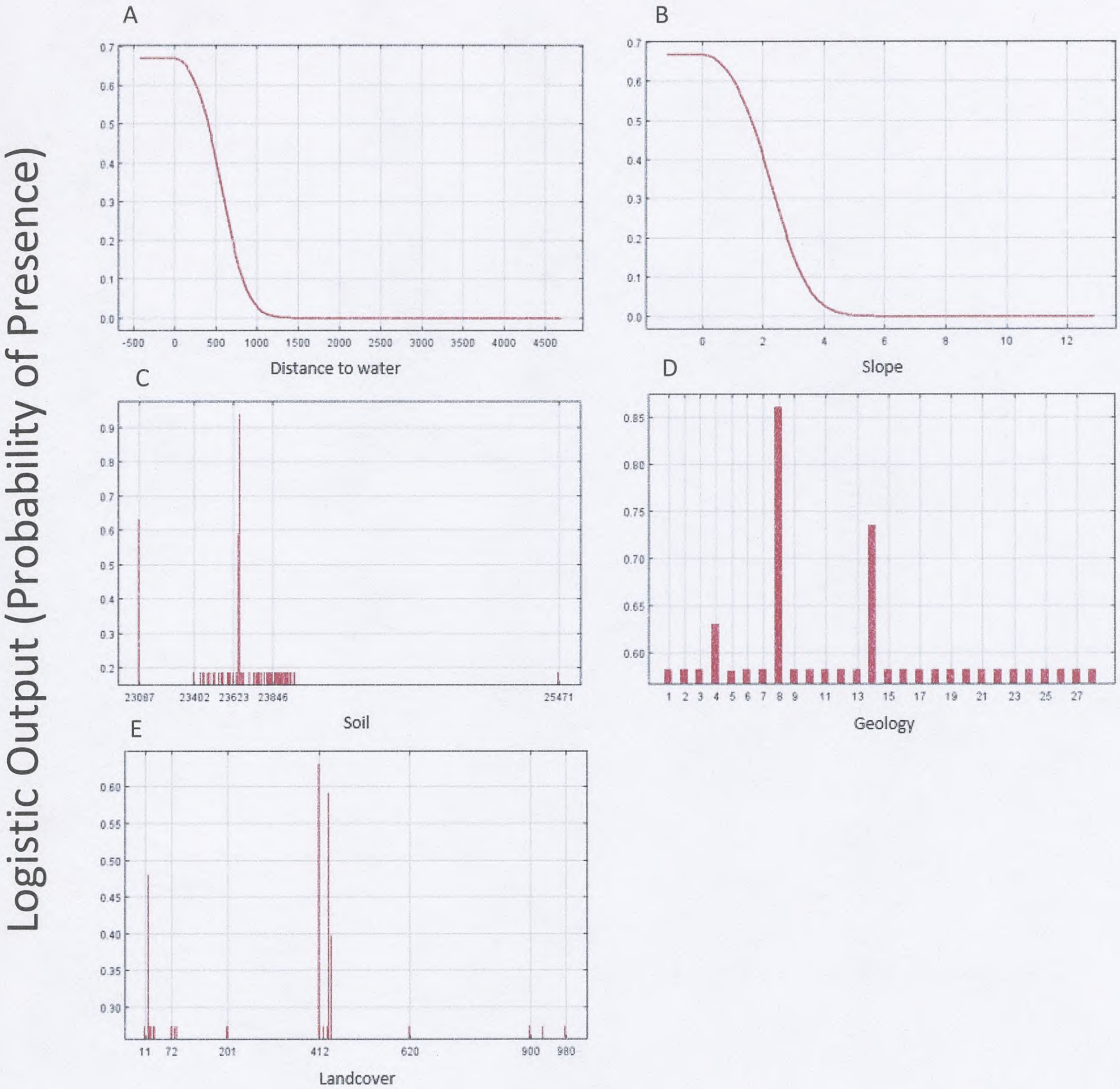


Figure 7. Species response for A) distance to water, B) slope, C) soil, D) geology, E) landcover. The Y axis symbolizes the probability of occurrence and the X axis depicts the environmental conditions (Appendix D). Negative values shown in A and B are model extrapolations but aren't used in the underlying model.

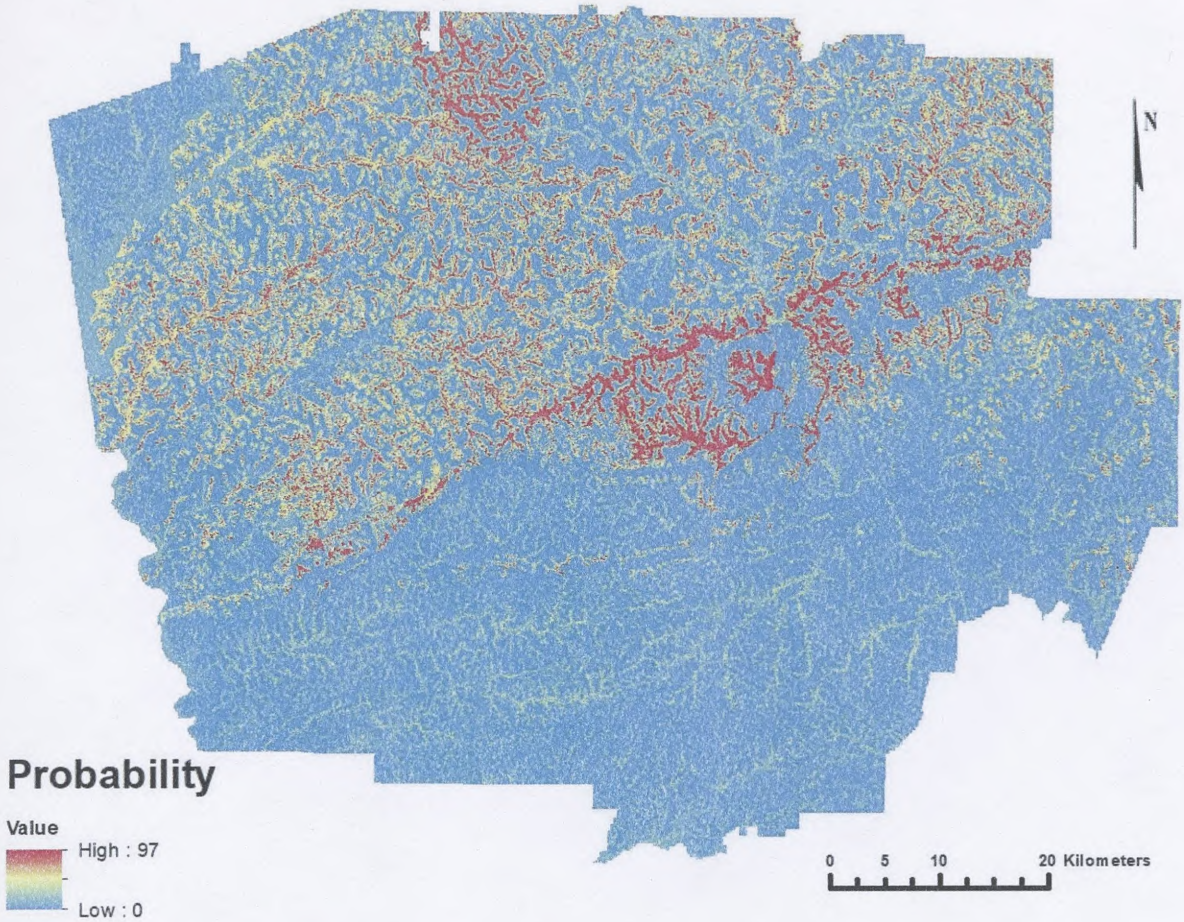


Figure 8. Maxent model predictions for the six counties analyzed in this study. Cool colors (ex: Blue/Whites) are areas of low probability of occurrence predictions, while the warmer (ex: Red) the color, the greater the probability of occurrence.

Table 2. The environmental layers and their contribution to the MaxEnt model. The permutation importance is the layer's correlation with the species while the percent contribution is the model assigned weight.

Variable	Percent contribution	Permutation importance
Distance to Water	35.4	29.2
Soil	29.1	27.1
Landcover	14.8	11.8
Geology	14.3	2.8
Slope	6.3	29.1

Field Validation: Random Locations

Sampling of the three different thresholds, low, medium, and high, resulted in no confirmed captures of *C. harti.* Photos illustrated differences in the herbaceous community among the three different thresholds (Appendix D). Wetland plants (i.e. arrowhead plants and ferns) were present at many locations of high model predicted probabilities whereas where the probability dropped the wetland vegetation became rare. The average time spent sampling locations increased from 40.7 min at low probability sites to 49.3 min at high probability sites (Table 3). The extended sampling effort was due a greater number of potential burrows dug as the quality of the habitat increased (Chi-square, $P < 0.01$, Table 3).

Table 3: An overview of the field validation results from sampling.

Sampling Probability	Average Time Spent	Standard Deviation	Potential Borrows Dug
Low (0%-20%)	40.7min	3min	0
Medium (40%-60%)	44.1min	6.8min	2
High (80%-100%)	49.3min	10.4min	10

Field Validation: Known Locations

To assess the model's predictive capacity at local scales, this study compared the probability scores at 3 sites with and without *C. harti* burrows. Counter to expectations model probabilities where crayfish weren't observed ranked higher than where they were observed (Fig. 9). Absence sites showed 20% higher probability scores (ANOVA, $P<0.001$, Table 4) than sites with *C. harti* present (Fig. 9). FDR Institute was removed from the ANOVA analysis, because all values at that site scored 0% probability. There existed statistical difference among sites (ANOVA, $P<0.001$, Table 4). The largest difference existed between Cartwright and Warm Springs ($P<0.001$) and the smallest between Chandler and Warm Springs (Least Squares Means, $P>0.14$). There was a significant interaction term present in the data (ANOVA, $P<0.001$, Table 4), because there was no significant difference between presence and absence at the Chandler and Cartwright locations (Least Squares Means, $P>0.57$, Fig. 10) while there was significant difference at Warm Springs (Least Squares Means, $P<0.001$, Fig. 10).

Table 4. Two-way ANOVA comparing model probabilities between presence/absence and sampling location (i.e. property).

Factor	Sum of Squares	df	Mean Square	F	p
Presence	2078.2	1	2078.2	23.289	< 0.001
Property	17662.3	2	8831.13	98.964	< 0.001
Presence * Property	807.1	2	403.57	4.523	0.016
Residual	4283.3	48	89.24		

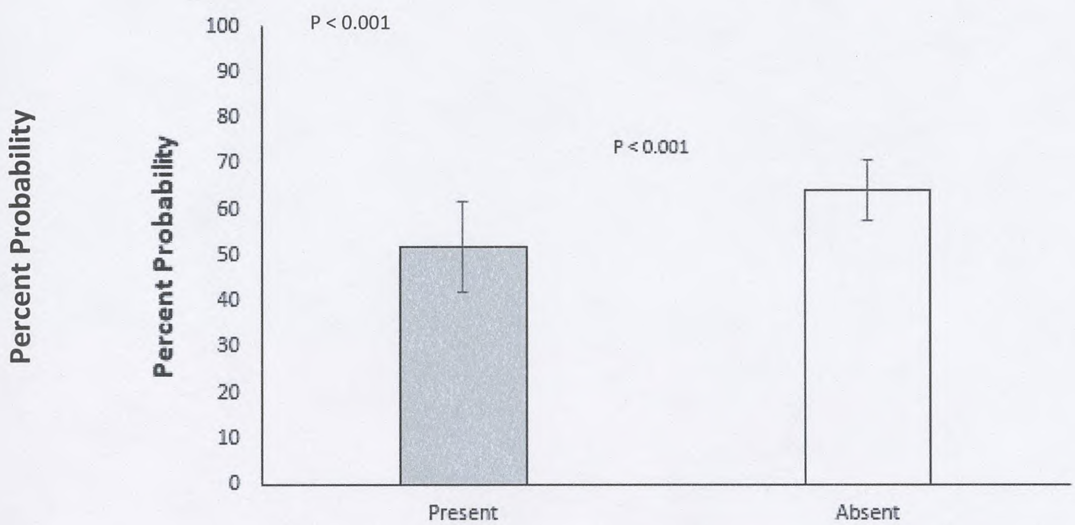


Figure 9. Combined average predicted model percent probabilities from 3 sites with presence (n=36) and absence (n=36) of *Cambarus harti.* Error bars represent 95% confidence intervals. The reported P-value is based on a two-way ANOVA.

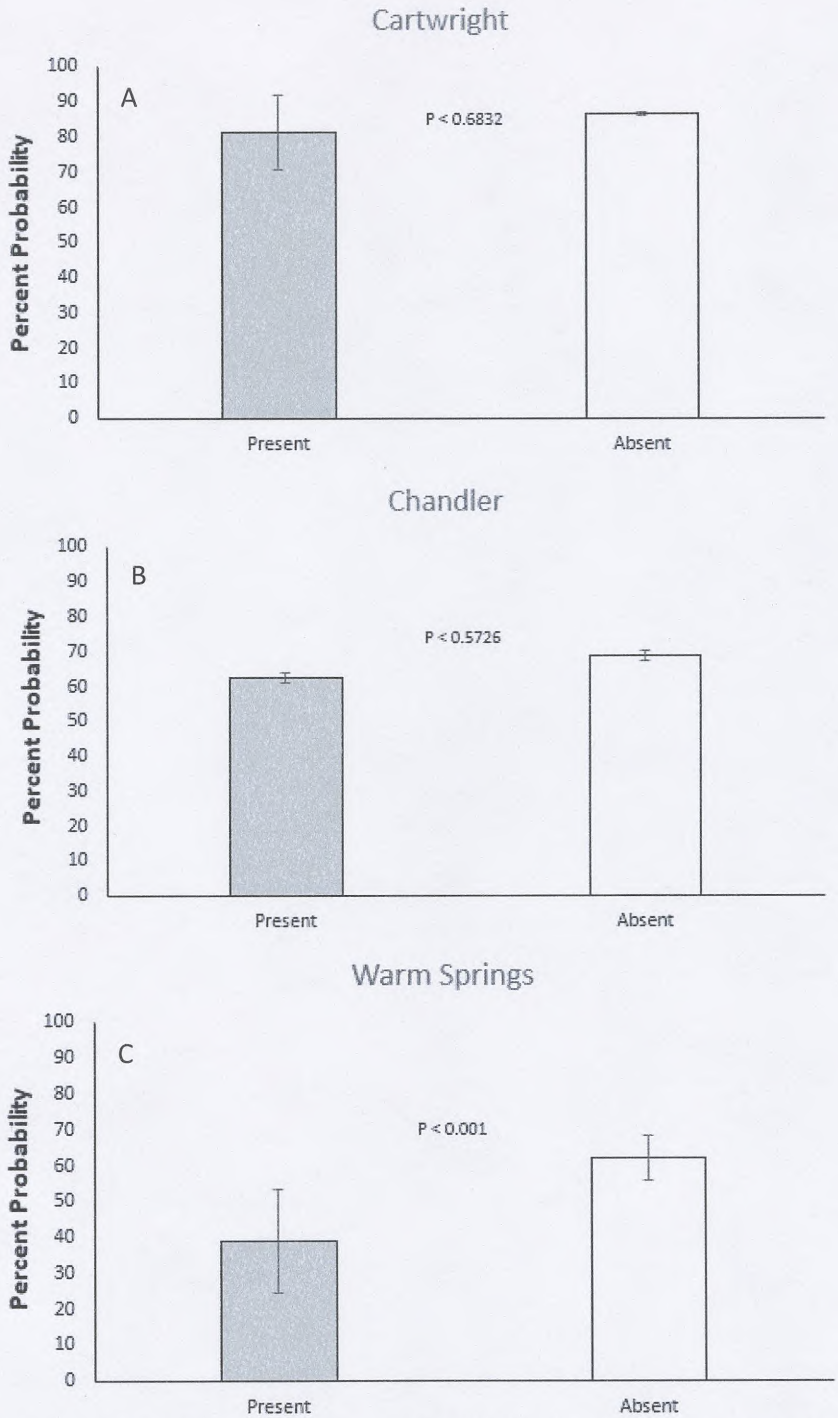


Figure 10. Predicted average model percent probabilities from A) Cartwright, B) Chandler, C) Warm Springs properties with and without *Cambarus harti.* Nine presence and nine absence probabilities were taken at each of the 3 sites. Error bars represent 95% confidence intervals. P-values were obtained from least squares means tests.

Discussion

MaxEnt has proven to be a powerful tool for expanding the known distributions of endangered species (Williams et al., 2009). This modeling approach was used in the Maryland Coastal Plain to predict sites for the endangered dwarf wedgemussel (Campbell & Hilderbrand, 2016) and endemic birds in temperate forests of Southern Chile (Moreno et al., 2011). MaxEnt has been successful when developing SDMs in cases where there are small sample sizes of only occurrence data (Hernandez et al., 2006, Wisz et al., 2008). Furthermore, it has already been used to model burrowing crayfish, *Fallicambarus harpi* and *Procambarus reimeri* in Arkansas (Rhoden et al., 2017). Given that *C. harti* was only known from 14 locations, MaxEnt was the appropriate species distribution model for this species. However, the results of this research appears to contradict previous publications regarding the effectiveness of MaxEnt for modeling rare burrowing crayfish (e.g. Rhoden et al., 2017).

The first form of ground validation for this model was conducted across thirty random locations, covering nearly 270 different model pixels. This intensive sampling resulted in no new occurrences of *C. harti,* not even in the high probability sites. There are several possible explanations for these findings. Previous research done on *C. harti* as well as other endangered burrowers indicated the potential for detection problems (Hobbs 1981). Hobbs (1981) reported that these crayfish are so limited in their distribution that it would require extensive sampling of all microhabitats within one location to confirm their presence (Hobbs, 1981). Time constraints allowed a total sample area of 83.54 square meters at each of the 30 validation sites: 10 high probability (100%-80%), 10 medium probability (40%-60%), and 10 low probability (0%-20%).

There exists the possibility that crayfish burrows were located in close proximity but were undetected during visual surveys. At three sites (11, 18, 30) with probabilities above 50%, vegetation such as ferns and broadleaf arrowhead indicated wetland conditions and likely the presence of shallow ground water (Kane et al, 2002). These indicator species are common at sites where this species is known to occur (Hobbs, 1981). The concern that a crayfish site existed just beyond the sampling locations was confirmed when a newly discovered location (32.884913$^{\circ N}$, -84.69916$^{\circ W}$), was reported 180m upstream from ground truthing site 14 (Appendix A). Not finding *C. harti* at any of the sites could indicate potential problems with the model fit (Tomarken & Waller, 2003) or the patchy distribution of rare species (Hobbs, 1981).

There exists potential for inaccuracies in SDM models even with high AUC's. MaxEnt determined the permutation of importance for each layer; distance to water, soil, geology, landcover, and slope (Table 2). From this the model compared how abundant these environmental conditions were across the landscape and determined the level of contribution that each layer would be weighted in the final model (Table 2). The model itself scored an AUC value of 0.957. This means that 95.7% of the time a random pixel chosen will score lower than one where the species is known to occur. This indicates that the habitat at the known locations were unique compared to the surrounding areas. Thus endangered endemic species with particular habitat requirements would be expected to have high AUCs (Rhoden et al., 2017). Researchers have identified problems with AUC values as indicators of model fit. A high AUC value could result from the model's limited capacity to estimate the habitat requirements (Warren & Seifert, 2011). This limited capacity could result from small sample sizes resulting in

a strong, but spurious correlation. Additionally, there exists potential for predicted absences and pseudo-absences (i.e. background points) to artificially inflate the AUC (Lobo et al., 2008). In this study, these potential pitfalls were taken into consideration by selecting adequate background points that were within 2km of each known location by following the recommendations of Peterman et al. (2013).

Because questions remained unanswered regarding the quality of the model at the local scale, a second phase of ground truthing was implemented. This approach examined the model on a larger scale (i.e. finer resolution) by comparing model probabilities at 4 locations where burrows were known to occur versus nearby sites where no burrows existed. While a significant difference was found between the model probabilities of these two different areas; surprisingly, locations without *C. harti* outscored the locations with *C. harti* (Tables 4 and 5). These findings proved that the model was not able to accurately predict burrows at a fine resolution. The limited sample size *C. harti* is something that must be taken into consideration. Studies on SDM's, such as the one by Hernandez (2006), indicate that MaxEnt preforms the best with limited sample sizes. However, at these small sample sizes the prediction success for the model could be as low as 20%. Wisz et.al (2008) also reported that MaxEnt preforms the best with limited sample sizes however at low sample sizes there exists a lower chance of prediction success.

One possible explanation for the model's poor performance, is that the known locations used for it weren't accurate. Site inaccuracies have been a problem for studies that extract occurrences from the literature (Newbold, 2010). However, all the sites used in the present study had been recently confirmed (Keller et al, 2011). This source of error seems unlikely to explain this model's poor performance.

Another source of error in the model could be traced to the environmental layers used to make the model itself. One study addressed the concern about standards for data collection and called for rigorous quality control plan for spatial data (Cayuela et al., 2009). Standards should be set to ensure that collection technique and data reporting form help ensure an adequate level of precision. Other Inaccuracies can cause the model to make false predictions particularly when there is lack of data and/or gaps in the existing data (Araujo & Guisan, 2006, Guisan & Thuiller, 2005). Gaps in data layers and data quality may have contributed to potential model inaccuracies identified by ground truthing in this study.

To determine if data layers contributed to the model prediction errors, the layers were examined visually in ArcMap. Locations where the species was known to exist were compared to each of the environmental layers and cross referenced with field observations. At all locations, it was observed that surface water from the USGS National Hydrology Dataset (NHD) showed inaccuracies (Appendix B, Image1-4). For example, at the Warm Springs property the NHD is missing a spring upwelling that forms a small stream (Appendix B, image 3). The known *C. harti* location was located in an area adjacent to this spring that was not identified in the NHD. In another case, the FDR Institute site scored a probability value of 0%, because the environmental layers inaccurately depicted the hydrography of the location (Appendix B, Image 4). At that site, there are crayfish living on both sides of a small spring-fed stream. According to the National Hydrology Dataset this stream, doesn't exist or hasn't been recorded. These errors suggest that the hydrography dataset contained important inaccuracies. Unfortunately, those data sets were used to formulate the model. Errors associated with missing or misplaced streams, contributed to an inaccurate final MaxEnt model.

In the future, the stream data should be revised and corrected especially if it will be used for another SDM. The model is only as good as the data used to create it. When data sets are lacking or inaccurate, the model will be unable create reliable predictions.

For burrowing crayfish models, an environmental layer that could improve model prediction inaccuracies in the NHD layer would be a layer depicting shallow groundwater. To date, there exists no groundwater layer for this area in Georgia. Considering that *C. harti* as well as other burrowing crayfish rely on groundwater connectivity, a layer like this would help create a more accurate species distribution model. Rhoden et al (2017) also lacked a groundwater layer, however they constructed a simple groundwater indicator. Their ground truth sampling did find the target species. However, they failed to properly test their model because they chose to sample only in roadside ditches where they knew water was present. By only sampling in ditches, the known preferred habitat of their crayfish, they may have biased the ground truthing and their estimate of the model's accuracy. In order to accurately ground truth the model the technique would need to be revised and sample the entire model landscape, rather than a selected subset of locations more likely to sustain the species.

Conclusion

SDM models have proven effective when modeling rare species, including burrowing crayfish (Rhoden et al., 2017). The quality of environmental layers and their relevance to the species being modeled, controls the accuracy of the model predictions. In this study the predicted distribution consisted of 6110km^2 with probabilities of occurrence ranging from: 0%-

100%. The resulting MaxEnt model showed significant inaccuracies in its prediction. One identified problem was that the environmental layers used in the model contained errors. For example analysis revealed that the USGS National Hydrology Dataset was missing springs and small streams, critical habitat for *C. harti*. These problems combined with the limited number of known locations contributed to the model's inaccurate predictions. Furthermore, critical environmental layers, such as surficial groundwater, does not exist for this area. These data are needed to accurately depict *C. harti* habitat. Until the data for the environmental layers are more fully developed and properly ground truthed, the value of SDM's for modeling rare burrowing crayfish will be limited.

Literature Cited

Araujo, M. B., & Guisan, A. (2006). Five (or so) challenges for species distribution modelling. Journal of Biogeography, 33(10), 1677-1688.

Aquatic Conservation Initiative | Georgia DNR - Wildlife Resources Division. (n.d.). Retrieved November 22, 2016, from http://www.georgiawildlife.com/node/2001/

Campbell, C. A., & Hilderbrand, R. H. (2016). Using maximum entropy to predict suitable habitat for the endangered dwarf wedgemussel in the Maryland Coastal Plain. Aquatic Conservation: Marine and Freshwater Ecosystems. doi:10.1002/aqc.2699

Cayuela, L., Golicher, D. J., Newton, A. C., Kolb, M., de Alburquerque, F. S., Arets, E. J. M. M. & Pérez, A. M. (2009). Species distribution modeling in the tropics: problems, potentialities, and the role of biological data for effective species conservation. Tropical Conservation Science, 2(3), 319-352.

Cordeiro, J., Skelton, C. & Thoma, R. F. (2010). *Cambarus harti*. IUCN 2010. IUCN Red List of Threatened Species.

Covich, A. P., Palmer, M. A., & Crowl, T. A. (1999). The Role of Benthic Invertebrate Species in Freshwater Ecosystems. BioScience, 49(2), 119-127. doi:10.2307/1313537

Crayfish of U.S. | Invertebrate Zoology Carnegie Museum of Natural History (n.d.). Retrieved October 25, 2016, from www.iz.carnegiemnh.org/crayfish/

Department Of Natural Resources Division. (n.d.). Retrieved October 21, 2017, from http://gadnr.org/

Dudík, M., Phillips, S. J., & Schapire, R. E. (2005). Correcting sample selection bias in maximum entropy density estimation. In Advances in neural information processing systems. Advances in Neural Information Processing Systems, (pp. 323-330).

Elith, J., Phillips, S. J., Hastie, T., Dudík, M., Chee, Y. E., & Yates, C. J. (2011). A statistical explanation of MaxEnt for ecologists. Diversity and Distributions, 17(1), 43-57. doi:10.1111/j.1472-4642.2010.00725.x

Gilmer, J.H. (2014). Characterizing Wetland Hydrology, Chemistry, and Soils for an Endemic Crayfish, the Piedmont Blue Burrower. MS Thesis, Columbus State University.

Gilmer J.H. and Keller T.A. (2015). Groundwater chemistry and soils have limited influence on the habitat-scale distribution of *Cambarus harti* Hobbs. Freshwater Crayfish 21. 43-50. 10.5869/fc.2015.v21-1.43.

Guisan, A., & Thuiller, W. (2005). Predicting species distribution: Offering more than simple habitat models. Ecology Letters, 8(9), 993-1009. doi:10.1111/j.1461-0248.2005.00792.x

Hasiotis, S.T. (1993). Evaluation of the burrowing behaviour of stream and pond dwelling species of *Orconectes* in the Front Range of Boulder, Colorado USA: Their ethological and geological implications. Freshwater Crayfish 9:399–406.

Hastie, T., Tibshirani, R. & Friedman, J.H. (2009) The elements of statistical learning: data mining, inference, and prediction, 2nd edn. Springer-Verlag, New York

Helms, B. S., Figiel, C., Rivera, J., Stoeckel, J., Stanton, G., & Keller, T. A. (2013). Life-history observations, environmental associations, and soil preferences of the Piedmont blue burrower (*Cambarus* [Depressicambarus] *harti*) Hobbs. Southeastern Naturalist, 12(1), 143-160.

Hernandez, P. A., Graham, C. H., Master, L. L., & Albert, D. L. (2006). The effect of sample size and species characteristics on performance of different species distribution modeling methods. Ecography, 29(5), 773-785.

Hobbs, H. H. (1981). The crayfishes of Georgia. Washington: Smithsonian Institution Press.

IBM Corp. Released 2017. IBM SPSS Statistics for Windows, Version 25.0. Armonk, NY: IBM Corp.

Kane, M. E., Philman, N. L., & Emshousen, C. (2002). Wetland plant propagation: Comparative growth and reproduction of micropropagated Sagittaria latifolia ecotypes. In Comb. Proc. Intl. Plant Prop. Soc (Vol. 52, pp. 454-459).

Keller, T.A., Rivera, J., Figiel, C., Stanton, G., (2011) "Ecology and biogeography of the Piedmont Blue Burrower (*Cambarus* [Depressicambarus] *harti*)" . CSU ePress.

Lobo, J. M., Jiménez-Valverde, A., & Real, R. (2008). AUC: a misleading measure of the performance of predictive distribution models. Global Ecology and Biogeography, 17(2), 145-151.

Loughman, Z.J. (2010). Ecology of *Cambarus dubius* (Upland burrowing crayfish) in north-central West Virginia. 2010. Southeastern Naturalist 9:217–230.

Master, L. L., Flack, S. R., & Stein, B. A. (Eds.). (1998). Rivers of life: critical watersheds for protecting freshwater biodiversity (p. 71). Arlington, Virginia: Nature Conservancy.

Moreno, R., Zamora, R., Molina, J. R., Vasquez, A., & Herrera, M. Á. (2011). Predictive modeling of microhabitats for endemic birds in South Chilean temperate forests using maximum entropy (Maxent). Ecological Informatics, 6(6), 364-370. doi:10.1016/j.ecoinf.2011.07.003

Newbold, T. (2010). Applications and limitations of museum data for conservation and ecology, with particular attention to species distribution models. Progress in Physical Geography, 34(1), 3-22.

Peterman, W. E., Crawford, J. A., & Kuhns, A. R. (2013). Using species distribution and occupancy modeling to guide survey efforts and assess species status. Journal for Nature Conservation, 21(2), 114-121.

Phillips, S. J. (2008). Transferability, sample selection bias and background data in presence-only modelling: A response to Peterson et al. (2007). Ecography, 31, 272–278.

Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2005). Maximum entropy modeling of species geographic distributions. Ecological Modelling, 190(3-4), 231-259. doi:10.1016/j.ecolmodel.2005.03.026

Phillips, S. J., Dudík, M., & Schapire, R. E. (2004). A maximum entropy approach to species distribution modeling. Twenty-first International Conference on Machine Learning - ICML '04. doi:10.1145/1015330.1015412

Phillips, S. J., & Dudík, M. (2008). Modeling of species distributions with Maxent: New extensions and a comprehensive evaluation. Ecography, 31, 161-175. doi:10.1111/j.0906-7590.2007.5203.x

Phillips S. J., Dudík M., Schapire R. E.. (2011) Maxent software for modeling species niches and distributions (Version 3.3.3k). Available from url: http://biodiversityinformatics.amnh.org/open_source/maxent/. Accessed on 2017-11-4.

Primack, R. B. (2006). Essentials of Conservation Biology, 4th ed. Sinauer Associates, Inc

R Core Team (2017). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org/.

Raes, N., & Steege, H. T. (2007). A null-model for significance testing of presence-only species distribution models. Ecography, 30(5), 727-736. doi:10.1111/j.2007.0906-7590.05041.x

Rhoden, C. M., Peterman, W. E., & Taylor, C. A. (2017). Maxent-directed field surveys identify new populations of narrowly endemic habitat specialists. PeerJ, 5, e3632.

Richman, N. I., Böhm, M., Adams, S. B., Alvarez, F., Bergey, E. A., Bunn, J. J. & Dawkins, K. L. (2015). Multiple drivers of decline in the global status of freshwater crayfish (*Decapoda*: *Astacidea*). Phil. Trans. R. Soc. B, 370(1662), 20140060.

Schläpfer, F. (1999). Expert estimates about effects of biodiversity on ecosystem processes and services. Oikos, 346-352.

Singh, J. S. (2002). The biodiversity crisis: a multifaceted review. Current Science, 82(6), 638-647.

Skelton, C.E. (2010). History, status, and conservation of Georgia crayfishes. Southeastern Naturalist 9:127–138.

Skelton, C., S. Cammack, & E. Van de Genachte. (2002). Surveys of rare burrowing crayfish species. Final report submitted to Georgia Natural Heritage Program, Wildlife Resources Division, Social Circle, GA.

Strayer, D. L., & Dudgeon, D. (2010). Freshwater biodiversity conservation: recent progress and future challenges. Journal of the North American Benthological Society, 29(1), 344-358.

Tarr, R. S. (1884). Habits of the burrowing crayfishes in the United States. Nature 30:127–128.

Taylor, C. A., Schuster, G. A., Cooper, J. E., DiStefano, R. J., Eversole, A. G., Hamr, P. & Thoma, R. F. (2007). A reassessment of the conservation status of crayfishes of the United States and Canada after 10+ years of increased awareness. Fisheries, 32(8), 372-389.

Tomarken, A. J., & Waller, N. G. (2003). Potential problems with" well fitting" models. Journal of Abnormal Psychology, 112(4), 578.

Wallace, J. B., & Webster, J. R. (1996). The role of macroinvertebrates in stream ecosystem function. Annual Review of Entomology, 41(1), 115-139.

Warren, D. L., Glor, R. E., & Turelli, M. (2010). ENMTools: A toolbox for comparative studies of environmental niche models. Ecography, 33(3), 607-611. doi:10.1111/j.1600-0587.2009.06142.x

Warren, D. L., & Seifert, S. N. (2011). Ecological niche modeling in Maxent: the importance of model complexity and the performance of model selection criteria. Ecological Applications, 21(2), 335-342.

Williams, J. N., Seo, C., Thorne, J., Nelson, J. K., Erwin, S., O'Brien, J. M., & Schwartz, M. W. (2009). Using species distribution models to predict new occurrences for rare plants. Diversity and Distributions, 15(4), 565-576.

Wilson, C. D., Roberts, D., & Reid, N. (2011). Applying species distribution modelling to identify areas of high conservation value for endangered species: A case study using Margaritifera (L.). Biological Conservation, 144(2), 821-829. doi:10.1016/j.biocon.2010.11.014

Wisz, M. S., Hijmans, R. J., Li, J., Peterson, A. T., Graham, C. H., & Guisan, A. (2008). Effects of sample size on the performance of species distribution models. Diversity and Distributions, 14(5), 763-773. doi:10.1111/j.1472-4642.2008.00482.x

Zavaleta, E. S., Pasari, J. R., Hulvey, K. B., & Tilman, G. D. (2010). Sustaining multiple ecosystem functions in grassland communities requires higher biodiversity. Proceedings of the National Academy of Sciences, 107(4), 1443-1446.

APPENDIX A

Semi-rendom sample location 14 is plotted on a map overlay along with a realitivly new location discovered by Chester Figiel. The newly discovered location was not far from the random sample that resulted in a non-presence finding.

APPENDIX B

Image 1. The Cartwright property with the NHD hydrology layer imposed on an image (blue Line). There is a stream to the east of this point (red Line) however the NHD map didn't include the smaller creek.

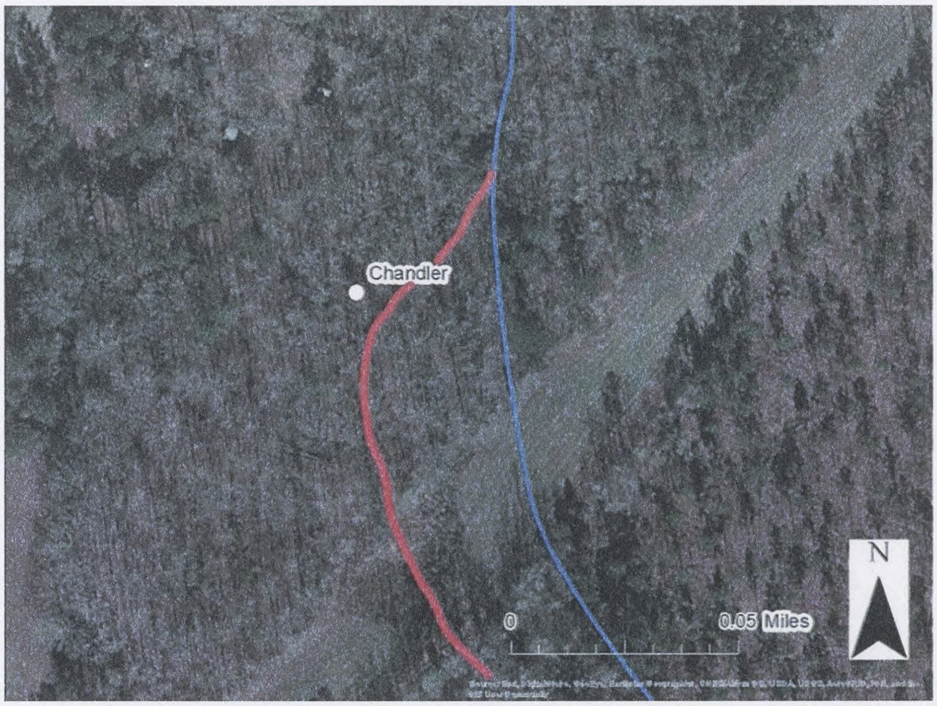


Image 2. The Chandler property with the NHD hydrology layer imposed on an image (blue line). The stream is drawn to the NE of my sample location however the streams true location (red line) runs parellel to my sample site before taking a bend and entering the easment.

Image 3. The Warm Springs property with the NHD hydrology layer imposed on an image (blue Line). The image depicts a stream to the east of the sample location however there is also a spring head (red dot) located at the sampling point and water from it flows towards the stream (red outline) before infiltrating into the ground.

Image 4. The FDR Institute property with the NHD hydrology layer imposed on an image (blue line). This image doesn't depict any water at this location, while sampling the point on the map a stream was evident (red line). The stream ran from up the hill towards my point and then under the road through a constructed stream crossing culvert.

Appendix C

Difference of Least Squares Means

Property	Presence	Property	Presence	Standard Error	DF	T Value	P Value
Cartwright		Chandler		3.1488	48	7.96	<.0001
Cartwright		Warm_Springs		3.1488	48	14.03	<.0001
Chandler		Warm_Springs		3.1488	48	6.07	<.0001
	Absent		Present	2.5710	48	4.83	<.0001
Cartwright	Absent	Cartwright	Present	4.4531	48	1.47	0.6832
Cartwright	Absent	Chandler	Absent	4.4531	48	5.54	<.0001
Cartwright	Absent	Chandler	Present	4.4531	48	7.19	<.0001
Cartwright	Absent	Warm_Springs	Absent	4.4531	48	8.03	<.0001
Cartwright	Absent	Warm_Springs	Present	4.4531	48	13.27	<.0001
Cartwright	Present	Chandler	Absent	4.4531	48	4.07	0.0023
Cartwright	Present	Chandler	Present	4.4531	48	5.71	<.0001
Cartwright	Present	Warm_Springs	Absent	4.4531	48	6.56	<.0001
Cartwright	Present	Warm_Springs	Present	4.4531	48	11.80	<.0001
Chandler	Absent	Chandler	Present	4.4531	48	1.65	0.5726
Chandler	Absent	Warm_Springs	Absent	4.4531	48	2.50	0.1459
Chandler	Absent	Warm_Springs	Present	4.4531	48	7.73	<.0001
Chandler	Present	Warm_Springs	Absent	4.4531	48	0.85	0.9566
Chandler	Present	Warm_Springs	Present	4.4531	48	6.09	<.0001
Warm_Springs	Absent	Warm_Springs	Present	4.4531	48	5.24	<.0001

APPENDIX D

Response curve tables

Table 1. Geology Dataset Attributes

Value	Abbreviation	Age	Rock Type 1	Rock Type 2
1	fg1	Precambrian-Paleozoic	Biotite Gneiss	Felsic Gneiss
2	bg1	Precambrian-Paleozoic	Biotite Gneiss	N/A
3	fg3	Precambrian-Paleozoic	Biotite Gneiss	Mica Schist
4	pms3a	Precambrian-Paleozoic	Mica Schist	Gneiss
5	mm3	Precambrian-Paleozoic	Gneiss	Amphibolite
6	fg4	Precambrian-Paleozoic	Biotite Gneiss	Amphibolite
7	gr1	Precambrian-Paleozoic	Granite	N/A
8	pms1	Precambrian-Paleozoic	Mica schist	N/A
9	gg6	Precambrian-Paleozoic	Granitic Gneiss	Granite
10	gg1	Precambrian-Paleozoic	Granitic Gneiss	N/A
11	pa2	Precambrian-Paleozoic	Schist	N/A
12	mm1	Precambrian-Paleozoic	Amphibolite	N/A
13	mm2	Precambrian-Paleozoic	Gneiss	N/A
14	q1	Precambrian-Paleozoic	Quartzite	N/A
15	water	Holocene	Water	N/A
16	c1	Age not given	Mylonite	N/A
17	q1a	Precambrian-Paleozoic	Quartzite	Mica schist

18	gr4	Precambrian-Paleozoic	Charnockite	N/A
19	pa1	Precambrian-Paleozoic	Schist	N/A
20	pms3	Precambrian-Paleozoic	Mica schist	Gneiss
21	mm4	Precambrian-Paleozoic	Gneiss	Amphibolite
22	bg2	Age not given	Biotite Gneiss	Amphibolite
23	c2	Age not given	Mylonite	N/A
24	Kt	Cretaceous	Sand	Clay or mud
25	Ke	Cretaceous	Clay or mud	Sand
26	Kb	Cretaceous	Clay or mud	Sand
27	Kc	Cretaceous	Sand	Clay or mud
28	Qal	Quaternary	Alluvium	Alluvial terrace

Table 2. Landcover Dataset Attributes

Value	Landcover Type	Description
7	Beach	Open sand, sandbars, mud and some sand dunes - natural environments as well as exposed sand from dredging and other activities. Mainly in coastal areas, but also inland, especially along the banks of reservoirs.
11	Open Water	Lakes, rivers, ponds, ocean, industrial water and aquaculture.
18	Transportation	Roads, railroads, airports and runways.
20	Utility Swaths	Open swaths maintained for transmission lines.

22	Low Intensity Urban - Nonforested	Low intensity urban areas with little or no tree canopy.
24	High Intensity Urban	Commercial/industrial and multi-family residential areas.
31	Clearcut - Sparse Vegetation	Recent clearcuts, sparse vegetation and other early successional areas.
33	Quarries, Strip Mines	Exposed rock and soil from industrial uses, gravel pits and landfills.
72	Parks, Recreation	Cemeteries, playing fields, campus-like institutions, parks and schools.
73	Golf Course	Golf courses.
80	Pasture, Hay	Pasture and non-tilled grasses.
83	Row Crop	Row crops, orchards, vineyards, groves and horticultural businesses.
201	Forested Urban - Deciduous	Low intensity urban areas containing mainly deciduous trees.
202	Forested Urban - Evergreen	Low intensity urban areas containing mainly evergreen trees.
203	Forested Urban - Mixed	Low intensity urban areas containing mixed deciduous and evergreen trees.
412	Hardwood Forest	Mesic to moderately mesic forests of the lower Piedmont and Coastal Plain. Includes non-wetland floodplain forests of yellow-poplar and sweetgum, ravines of oaks and American beech, and many upland oak-hickory stands.
413	Xeric Hardwood	Dry hardwood forests found throughout the state, although most common in the mountain regions, and progressively more rare southward. Includes areas dominated by southern red oak, scarlet oak, post oak and blackjack oak.

422	Open Loblolly-Shortleaf Pine	Only mapped in the Piedmont. Includes older, fairly open stands that may be almost savanna-like in appearance.
432	Xeric Mixed Pine-Hardwood	Dry mixed forests found throughout the state, although most common in the mountain regions and progressively more rare southward. Includes areas dominated by a mix of pines (most frequently shortleaf or Virginia in the mountains, and shortleaf or longleaf elsewhere) and hardwood species such as southern red oak, scarlet oak, post oak and blackjack oak.
434	Mixed Pine-Hardwood	Mesic to moderately dry forests of mixed deciduous and evergreen species found throughout the state at lower elevations. May include areas dominated by sweetgum, yellow-poplar, various oak species and loblolly or shortleaf pine.
440	Loblolly-Shortleaf Pine	Found from the upper Coastal Plain northward (rare in the Blue Ridge except at the lowest elevations). Includes many stands heavily managed for silviculture as well as areas regenerating from old field conditions.
512	Sandhill	Areas of scrub vegetation on deep, sandy soils on the Coastal Plain, especially near the Fall Line and along larger streams. May be dominated by turkey oak, blackjack oak, live oak, holly and longleaf pine.
620	Longleaf Pine	Open, savanna-type stand. Heavily managed plantations would likely be classed with 440 or 441. Most common on the lower Coastal Plain, although found up to the lower Piedmont and historically in the Ridge and Valley.
890	Cypress-Gum Swamp	Regularly flooded swamp forests mainly found on the Coastal Plain. May include either riparian or depressional wetlands. Usually dominated by pond or baldcypress and/or tupelo gum.
900	Bottomland Hardwood	Less frequently flooded wetland forests found throughout the state, but most common on the Coastal Plain. To the north, may be dominated by sweetgum, elms and red maple. To the south, wetland oaks (water oak, willow oak, overcup oak, swamp chestnut oak), black gum, and even spruce pine become more common.

930	Freshwater Marsh	Emergent freshwater wetlands found throughout the state. May be dominated by grasses or sedges.
980	Shrub Wetland	Closed canopy, low stature woody wetland. Found throughout the state, although most common on the Coastal Plain. May be result of clearcutting of wetland forests. Frequently includes willows, alders and red maple.
990	Evergreen Forested Wetland	Restricted to the Coastal Plain. Includes forests dominated by bay species, wet pine forests (typically slash or pond pine) or Atlantic white cedar.

Table 3. Soil Dataset Attributes

Value	AWC	CLAY	KFFACT	OM	PERM	HYGRP	DRAIN	LL	IFHYDRIC	AFLDFREQ
0	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
1	0.11	35.1	0.22	0.3	1.96	2	3	43.2	0	3.7
2	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
3	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1	-0.1
4	0.13	19	0.22	0.7	3.11	2.6	4.2	28.1	0.1	2.2
5	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
6	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
7	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
8	0.13	19	0.22	0.7	3.11	2.6	4.2	28.1	0.1	2.2
9	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
10	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
11	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
12	0.13	19	0.22	0.7	3.11	2.6	4.2	28.1	0.1	2.2

13	0.11	35.1	0.22	0.3	1.96	2	3	43.2	0	3.7
14	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
15	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
16	0.1	23.1	0.3	0.6	2.56	3.1	3	30.6	0	4
17	0.1	23.1	0.3	0.6	2.56	3.1	3	30.6	0	4
18	0.09	31.8	0.18	0.2	1.57	2.3	3	41.7	0	3.8
19	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
20	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
21	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
22	0.1	23.1	0.3	0.6	2.56	3.1	3	30.6	0	4
23	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
24	0.1	23.1	0.3	0.6	2.56	3.1	3	30.6	0	4
25	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
26	0.1	23.1	0.3	0.6	2.56	3.1	3	30.6	0	4
27	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
28	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
29	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
30	0.1	23.1	0.3	0.6	2.56	3.1	3	30.6	0	4
31	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
32	0.1	23.1	0.3	0.6	2.56	3.1	3	30.6	0	4
33	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
34	0.09	31.8	0.18	0.2	1.57	2.3	3	41.7	0	3.8
35	0.1	23.1	0.3	0.6	2.56	3.1	3	30.6	0	4

36	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
37	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
38	0.08	28.9	0.14	0.2	1.22	2.7	3.1	39.6	0	3.7
39	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
40	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
41	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
42	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
43	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
44	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
45	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
46	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
47	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
48	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
49	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
50	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
51	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
52	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
53	0.08	28.9	0.14	0.2	1.22	2.7	3.1	39.6	0	3.7
54	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
55	0.1	23.1	0.3	0.6	2.56	3.1	3	30.6	0	4
56	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
57	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
58	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9

59	0.1	23.1	0.3	0.6	2.56	3.1	3	30.7	0	4
60	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
61	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
62	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
63	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
64	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
65	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
66	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
67	0.1	23.1	0.3	0.6	2.56	3.1	3	30.6	0	4
68	0.09	37.6	0.16	0.2	1.06	3.2	3.6	52.5	0	3.9
69	0.09	37.6	0.16	0.2	1.06	3.2	3.6	52.5	0	3.9
70	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
71	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
72	0.09	37.6	0.16	0.2	1.06	3.2	3.6	52.5	0	3.9
73	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
74	0.09	37.6	0.16	0.2	1.06	3.2	3.6	52.5	0	3.9
75	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
76	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
77	0.09	37.6	0.16	0.2	1.06	3.2	3.6	52.5	0	3.9
78	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
79	0.11	26.2	0.22	0.2	2.76	2.1	3	33.2	0	3.9
80	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
81	0.09	17.8	0.2	0.4	5.48	2.2	2.9	30	0.1	3.8

82	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
83	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
84	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
85	0.08	15.4	0.17	0.4	6.87	1.9	2.4	29.7	0	3.9
86	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
87	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
88	0.11	35.6	0.23	0.2	1.27	2	3	44.7	0	3.9
89	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
90	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
91	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
92	0.09	17.8	0.2	0.4	5.48	2.2	2.9	30	0.1	3.8
93	0.12	32.9	0.26	0.3	2	2.1	3.2	39.7	0	3.8
94	0.09	17.8	0.2	0.4	5.48	2.2	2.9	30	0.1	3.8
95	0.11	26.2	0.22	0.2	2.76	2.1	3	33.2	0	3.9
96	0.08	15.4	0.17	0.4	6.87	1.9	2.4	29.7	0	3.9
97	0.1	17.4	0.18	0.5	5.48	1.7	2.8	30.3	0	3.8

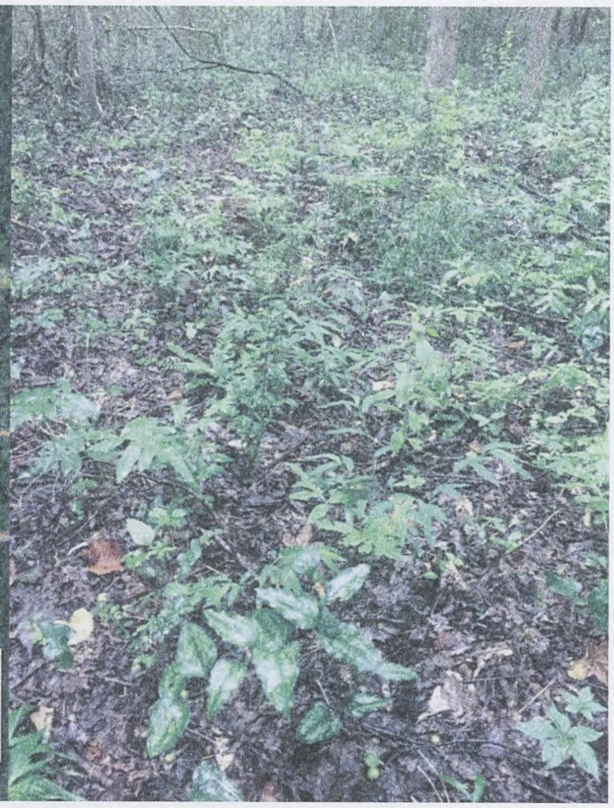
AWC = available water capacity (inches/inch)
CLAY = clay content of soil (% of soil < 2mm in size)
KFFACT = soil erodibility f-factor
OM = organic matter content (% by weight)
PERM = permeability rates (inches/hour)
HYGRP = soil index variables (1=well drained to 4=poorly drained)
DRAIN = soil index variable (1=well drained to 7=poorly drained)
LL = liquid limit of the soil (%moisture by weight)
IFHYDRIC = hydric soil indicator (1 if hydric)
AFLDFREQ = annual flood frequency (1 = frequent (>50% chance)
2 = occasional (5-50% chance), 3 = rare (<5% chance)

Appendix E

This is a compliation of many different site locations. Each set of pictures includes a location map depicting the area to help give reference followed by an image from the site that was taken in the centroid of my sample location. The series of images are split up into three different groups based on the probablity of occurrence at these locations (high, medium, low). The images depict a change in the groundcover as you move from high probablity to low. Within the higher probablities you will see groundcover that denotes shallow groundwater while the low probablities lack any cover and if there is some it doesn't indicate the presence of groundwater.

Locations of High Probability 80%-100%


Sample Location 13
N

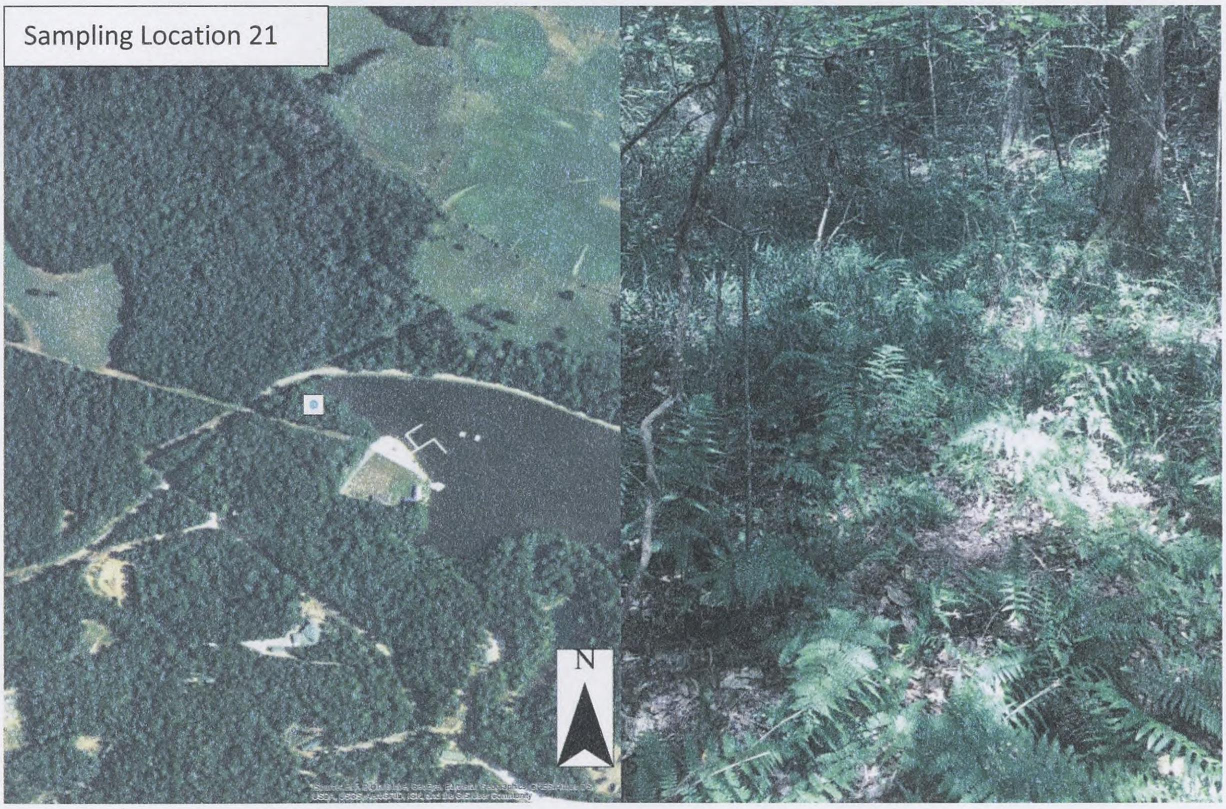



Sampling Location 14
N

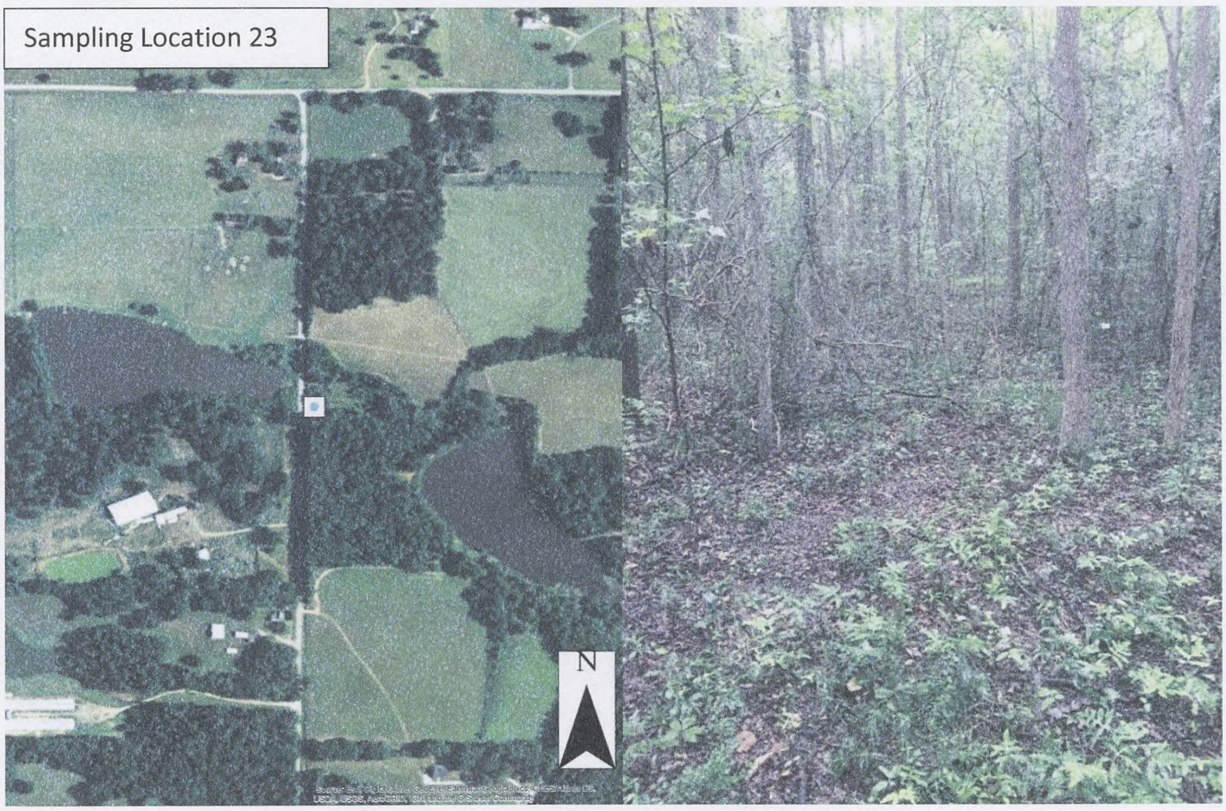

Sampling Location 16
N


Sampling Location 19
N


Sampling Location 20
N

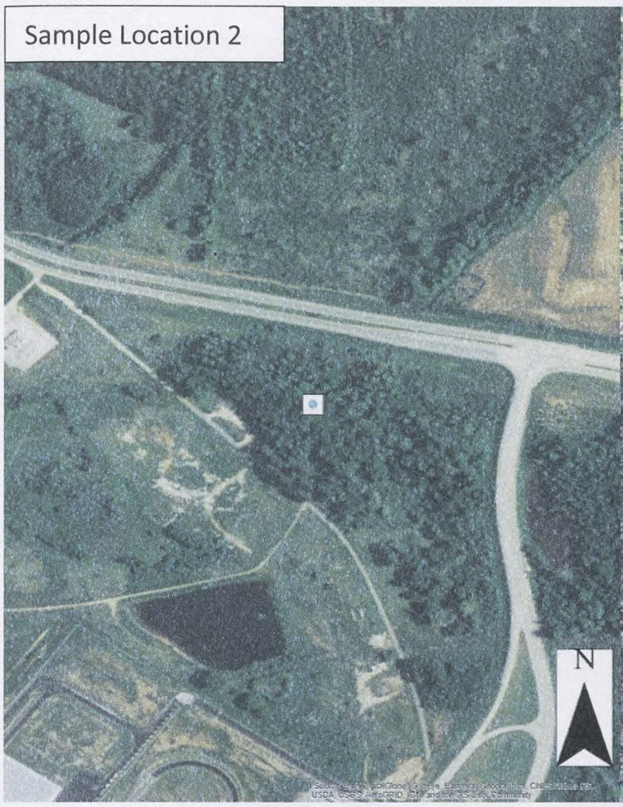

Sampling Location 21
N


Sampling Location 22
N


Sampling Location 23
N

Sampling Location 24

Locations of Medium probability 40%-60%




Sample Location 4
N


Sample Location 5
N


Sample Location 8
N


Sample Location 11
N

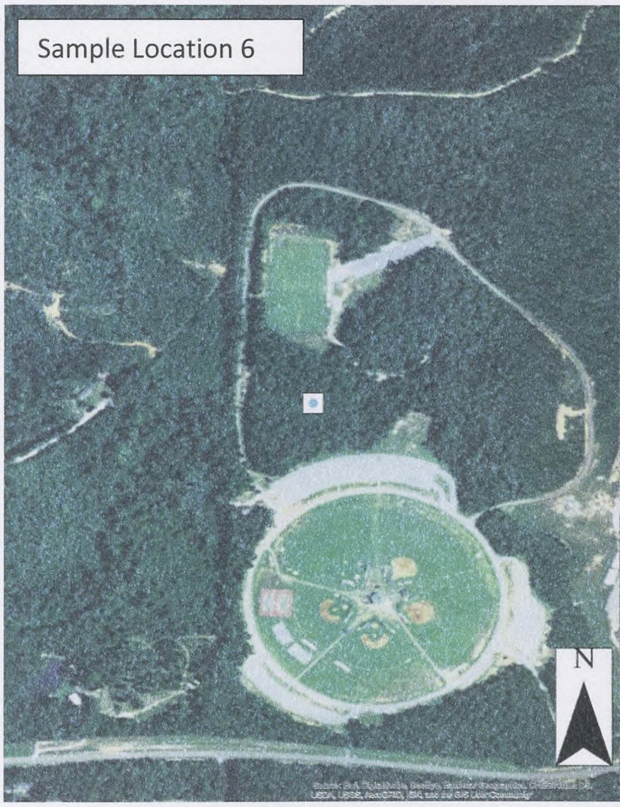
Sample Location 12

Sample Location 18

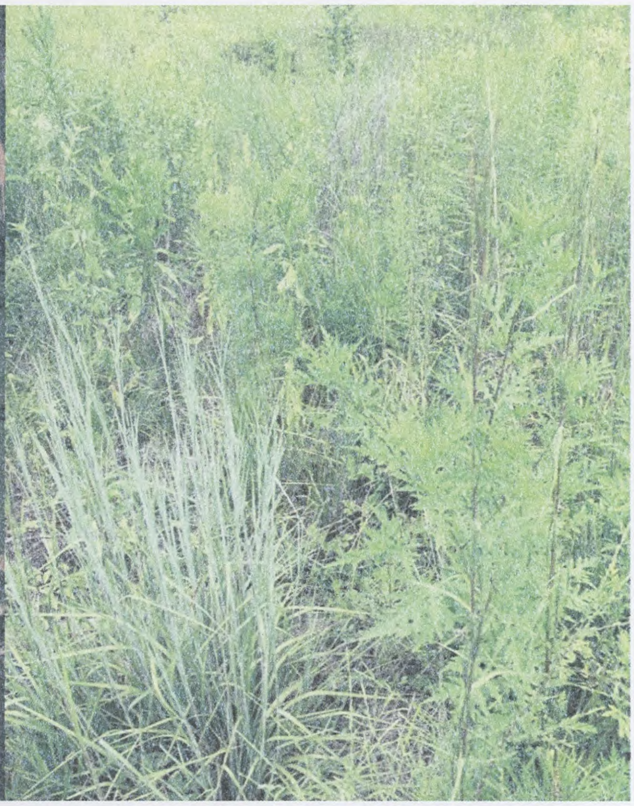
Sample Location 29

Sample Location 30

Locations of low probability 0%-20%



Sample Location 9



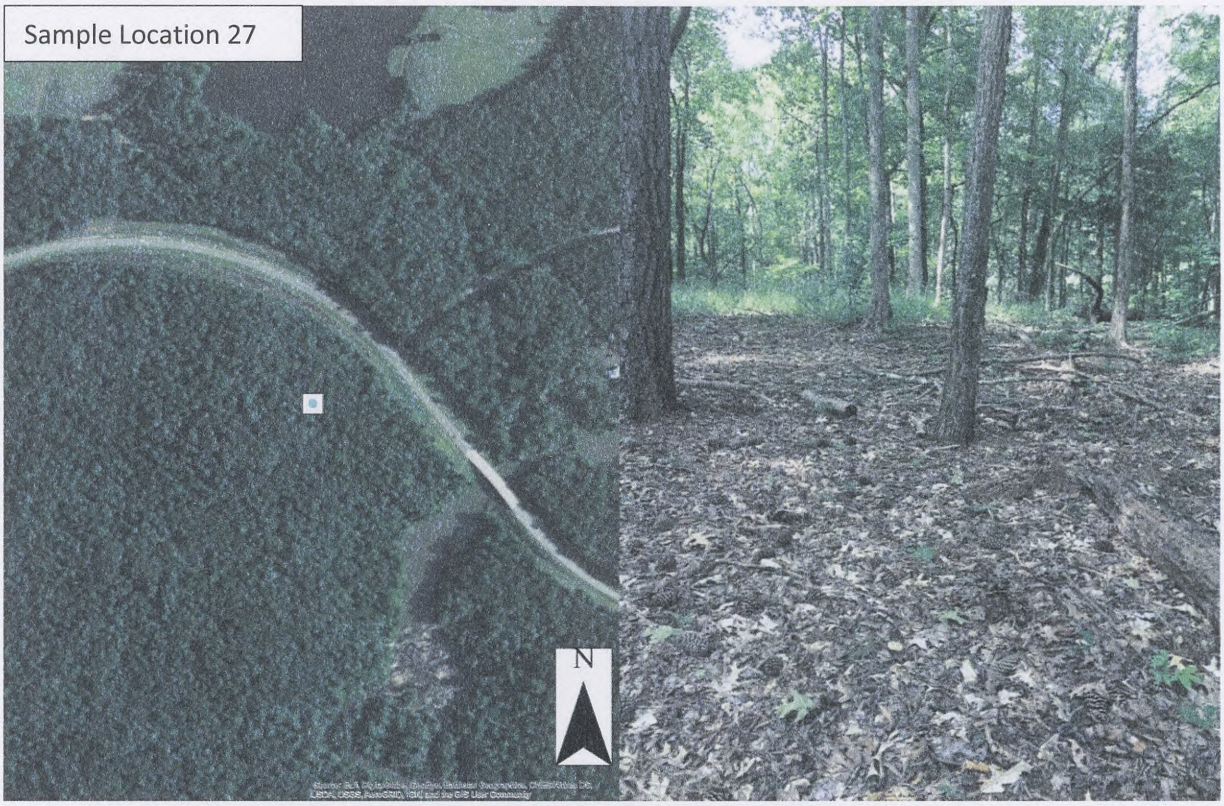
Sample Location 10

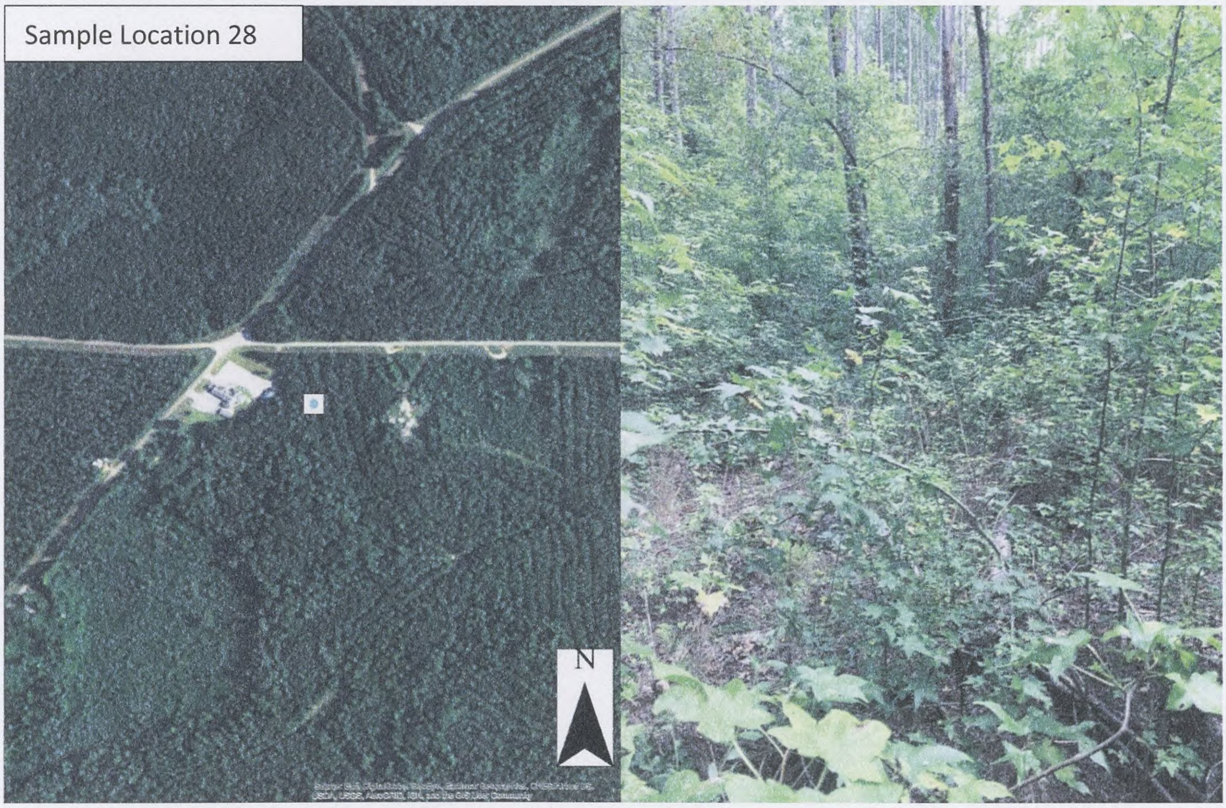

Sample Location 15
N


Sample Location 17
N


Sample Location 25
N


Sample Location 26
N


Sample Location 27
N


Sample Location 28
N

Appendix F

List of locations for all reandom sampled locations along with there probablity ranking.

probability	latitude	longitude
80-100	32.98941562	-84.47376288
80-100	32.85426801	-84.64408113
80-100	32.94832094	-84.46982345
80-100	33.00053076	-84.46486708
80-100	32.88720225	-84.69877335
80-100	32.96729112	-84.5024646
80-100	32.96441528	-84.54895304
80-100	32.82394499	-84.86487466
80-100	32.92182338	-84.65690954
80-100	32.95769792	-84.57873974
40-60	33.18857895	-84.6754586
40-60	33.17348199	-84.30272959
40-60	33.05023915	-84.32289425
40-60	32.83064132	-84.84297827
40-60	32.93013346	-85.1213369
40-60	32.77126974	-84.69577841
40-60	32.79563937	-84.57422811
40-60	32.7475528	-85.02140014
40-60	33.01889219	-85.04892564
40-60	32.99352661	-85.17156258
0-20	32.62275657	-84.70609441
0-20	32.65278022	-84.54150938
0-20	32.90541716	-84.36474301
0-20	32.98247257	-84.30420223
0-20	33.04760524	-84.70597371
0-20	32.67724191	-84.9253337
0-20	32.78258801	-84.40412751
0-20	32.92483811	-84.7709989
0-20	33.04640909	-84.51551196
0-20	32.91935118	-85.16524799